Promoters
Prof. dr. ir. P. Goethals
Laboratory of Aquatic Ecology
Department of Applied Ecology and Environmental Biology
Ghent University

Prof. dr. B. De Baets
Research Unit Knowledge-based Systems
Department of Applied Mathematics, Biometrics and Process Control
Ghent University

Dean
Prof. dr. ir. H. Van Langenhove

Rector
Prof. dr. P. Van Cauwenberge

ir. Ans Mouton

A critical analysis of performance criteria for the evaluation and optimisation of fuzzy models for species distribution

Thesis submitted in fulfilment of the requirements for the degree of
Doctor (PhD) in Applied Biological Sciences

Dutch translation of the title:
Een kritische analyse van performantiecriteria voor de evaluatie en optimalisatie van vage modellen voor soortenverspreiding

Please refer to this work as follows:
Mouton, A., 2008. A critical analysis of performance criteria for the evaluation and optimisation of fuzzy models for species distribution. PhD thesis, Ghent University, Gent, Belgium.

ISBN-number: 978-90-5989-248-4

This research was supported by a fellowship of the Special Research Fund (BOF) of Ghent University. Part of this research was supported by the *Leven met Water* programme (Water Framework Directive Explorer project) and by the *Delft Cluster* programme (Kaderrichtlijn Water Tools project).

Acknowledgements

'Writing a PhD is much like being a water drop in a large braided river: you get thrown into it, explore a wide range of directions and finally arrive at a destination you could never imagine at the start of your journey'.

De eerlijkheid gebiedt me te vertellen dat er onderweg ook enkele waterdruppeltjes zijn die de zee nooit halen: ze verdampen, infiltreren of worden domweg door een koe opgedronken. Met dit schrijven wil ik dan ook een aantal personen danken voor hun onmisbare hulp op weg naar zee.

Vooreerst dank aan mijn beide promotoren, Prof. Peter Goethals en Prof. Bernard De Baets. Peter, je hebt me steeds de kans gegeven om talrijke paden te ontdekken en zelf de meest aangewezen weg te kiezen. Bedankt voor je vertrouwen en voor de goede raad op cruciale momenten. Bernard, door je uitgebreide feedback heb je niet alleen de kwaliteit van dit werk aanzienlijk verbeterd, je hebt me ook de moed en het zelfvertrouwen gegeven om door te doen op moeilijker momentjes. Ik wil je ook bedanken om in de laatste weken ondanks alles toch tijd vrij te maken en hoop dat je er snel weer bovenop bent.

Verder wil ik Em. Prof. Niels De Pauw bedanken voor zijn steun tijdens mijn onderzoek. Zijn zin voor (taalkundige) correctheid en zijn kunst om van het leven te genieten zijn ook mij bijgebleven. Bovendien stelde hij ook altijd het menselijke aspect centraal, wat ik enorm apprecieer.

Ik wil de andere leden van de examencommissie, Prof. dr. ir. W. Bauwens, Prof. Dr. A. Vanreusel en Prof. dr. ir. W. Verstraete bedanken voor hun waardevolle commentaren en suggesties om dit werk te verbeteren.

A very warm thank you goes to Matthias Schneider and his family: Tine, Benny and Frederic. Your hospitality while I was in Stuttgart made me feel right at home every time and it was wonderful to be a part of the family. I really hope one day you can make this trip to Belgium. Matthias, thank you for the chances you gave me to work with CASiMiR and MesoCASiMiR and to learn so many other things. One of them is that, along the way to a PhD, you can meet warm and interesting people, and also make a friend for life.

I would like to express my gratitude also to Andreas Eisner, Franz Kerle, Gregory Egger, Iana Kopecki, Klaus Jorde, Kristin Eastman, Markus Noack and Silke Wieprecht for the fruitful discussions in many places around the world, and for training my German to become better than that of a famous Belgian goalkeeper.

A special thank you is for Ian Jowett and Sally, who were so friendly to host a complete stranger during his stay in New Zealand. Ian, thanks you very much for the opportunity to work with you at NIWA; the time was way too short to share all your experience but I definitely learned a lot. The fishing trips were a wonderfull weekend activity which I'll never forget! I hope you'll keep on catching big snappers and maybe some day we can try fishing Belgian rivers (not as spectacular though).

During the last 4 years, I also had the chance to join the European COST 626 programme, in which I met many interesting people. Thank you Armin Peter, Atle Harby, Dough Booker, Graham Hill, Harald Meixner, Harm Duel, Helmut Mader, Ian Maddock, Mike Dunbar, Monica Rivas Casado, Morten Lauge Pedersen, Nataša Smolar, Nicolai Friberg, Nicolas Lamouroux, Peter Borsányi, Piotr Parasiewicz, Sylvie Mérigoux and many others for interesting discussions, nice cooperation and good fun during various meetings throughout Europe.
In het bijzonder wil ik ook dr.(!) Bruno Verbist bedanken voor zijn grenzeloze gastvrijheid tijdens ons verblijf in Indonesië, voor het opvangen van de cultuurshock (AC/DC rules!) en voor de vele leuke gesprekken in meer en minder exotische oorden. Proficiat ook Bruno, je bent er in geslaagd om me op de meet te kloppen!

Aan alle collega's met wie ik de voorbije 4 jaar heb samengewerkt binnen AECO en KERMIT: een dikke merci! De redenen zijn te talrijk om op te sommen, maar ieder van jullie weet ongetwijfeld wel waarom. Bedankt An, Anke, Annelies M., Argaw, Bea, Bram, Diederik, Eleni, Gisèle, Guido, Hilde, Koen, Lien, Luis, Marianne, Marjolein, Martine, Minh, Rahmat, Spyros, Thu Huong, Wim, Wouter en Veronique. Andy, bedankt voor het fijne verblijf in Indonesië, toastbrood met hagelslag is sindsdien telkens weer een feest. Tom, jou wil ik graag bedanken voor de cruciale tips die je me aan het begin van de weg hebt gegeven, en voor de aanmoediging tijdens de steilere stukjes. Ester, het was enorm fijn samenwerken en redeneren met jou, hopelijk vinden we hier vanaf nu weer wat meer tijd voor. Verder ook bedankt aan alle thesisstudenten: Annelies, Collette, Elisa, Els, Esther, Johan, Matthew, Piet, Pol, Tim en Tine. Aan alle collega's van de ecotoxgroep: dank voor de leuke koffiepauzes, seminaries, sportnamiddagen en après-PhD exploten. Een bijzonder bedankje aan Michiel en Roel voor de hulp bij de receptie. Ook dank aan de collega's van het Centrum voor Milieusanering, het was altijd plezant bij jullie op het tweede.

Ik ben enorm blij dat ik nog steeds af en toe naar m'n 'roots' kan terugkeren. Bedankt vrienden van Sente voor de vele jaren waarin we samen opgroeiden en elkaar toch niet uit het oog verloren. Alexander, Bart, Bram, Ignace, Jan, Peter, Rando, Reinhilde, Rob, Sara, Sofie, Tim, Waldo en ondertussen al vele partners en andere jeugdhuismonumenten, de ontelbare kleine momentjes zijn onvergetelijk. Hopelijk zijn we nog maar net begonnen..
Bedankt ook aan de vrolijke vrienden die de Gentse stede met kleur doorspekten: Brecht, Dieter, Jochen, Joke, Katrien, Koen V., Koen V.M., Lieslotte (Oma!), Michael, Nathan, Wim, Wouter en vele partners, VVMT-ers en mede-studenten. Stef, Bram en Kobe, Sarah en Marijn, Anke, Damir, Ilena en ??, bedankt voor de vele fijne avondjes, Dranouters, Irish Coffees en andere activiteiten.

Mijn familie, nonkels, tantes, neven, nichten wil ik bedanken voor de goede samenhang doorheen de jaren; hopelijk mag dit nog lang zo blijven. Pépé, bedankt voor je eenvoud, je zijn en de kunst om te genieten van de kleine dingen in het leven. Marie-Jeanne en Armand, een hele grote 'dank u' om een warme thuis te zijn aan de andere kant van het land. Bedankt ook Marilou en Herwig, Peter, An, Marthe en Michiel om me zo hartelijk in de familie op te nemen.

'Het zijn niet de titels van een mens die het leven zinvol maken, het zijn de verdoken kleine eilanden waarop men zichzelf is geweest.'

Ma en pa, ik krijg kippenvel als ik denk aan alles wat jullie voor me hebben gedaan. Het is bijna onmogelijk om er slechts enkele dingen uit te pikken, maar ik wil jullie toch bedanken voor het grenzeloze vertrouwen en de niet aflatende steun in de voorbije 27 jaar.

Schatje... zonder jou was ik dit nu niet aan het schrijven. Er bestaan eigenlijk geen woorden om je te bedanken, maar ik probeer het toch. Bedankt voor jezelf, voor de waarde-volle, sterke vrouw die 3 jaar lang m'n steun en toeverlaat is geweest, voor het vertrouwen en voor het geduld; ik had mezelf waarschijnlijk al lang buitengezwierd. Dank je om me er af en toe aan te herinneren dat er leven is naast het doc, ik zie er naar uit om er weer van te genieten!

Contents

List of Abbreviations

aAD	Adjusted Average Deviation
AD	Average Deviation
AIC	Akaike Information Criterion
AUC	Area Under the (Receiver-Operator) Curve
AUSRIVAS	Australian River Assessment System
BEAST	Benthic Assessment of Sediment
CCI	Correctly Classified Instances
EU	European Union
HSI	Habitat Suitability Index
Kappa	Cohen's *Kappa*
NMI	Nominal Mutual Information Statistic
PHABSIM	Physical Habitat Simulation System
RIVPACS	River Invertebrate Prediction and Classification System
ROC	Receiver-Operator Characteristic
Sn	Sensitivity
Sp	Specificity
TSS	True Skill Statistic
WFD	Water Framework Directive
WUA	Weighted Usable Area
WUV	Weighted Usable Volume

CHAPTER 1

General introduction

1.1. Setting

For many years, river management in Flanders (Belgium) has been conducted at the basin level, using instruments such as wastewater treatment plants and enforced effluent standards. Although these measures resulted in a significant improvement of the chemical and ecological river quality, still about 35 % of the domestic wastewater in Flanders is being discharged in the surface waters without any treatment (De Cooman et al., 2007). To meet the aims set by the Water Framework Directive (EU, 2000), the Flemish government has planned to reduce this discharge to 20 % in 2007 (De Cooman et al., 2007). Moreover, small-scale measures such as remeandering, flood plane restoration and fish passages are planned or are being implemented. However, most of these restoration measures are being conducted without any quantitative prediction of their impact on river ecology. Consequently, efficient allocation of these efforts is needed. Since habitat suitability or species distribution models can detect the bottlenecks in a river basin, these models can support river managers in this allocation process and may therefore be valuable tools for integrated river management (Poff et al., 2003).

For the development of these decision support tools, there is today a growing interest in modelling techniques such as artificial neural networks (Lek and Guégan, 1999), decision trees (Džeroski, 2001) and fuzzy logic (Adriaenssens et al., 2004). To allow decision makers to assess the uncertainty associated with the model outputs, the model structure should be straightforwardly interpretable (Omlin and Reichert, 1999; Elith et al., 2002; Regan et al., 2002; Brugnach et al., 2003; Borsuk et al., 2004). Fuzzy logic has become an interesting technique to address this issue. It takes into account the inherent uncertainty of ecological variables during inference and it enables expressing non-linear relations between ecological variables in a transparent way (Salski, 1992; Silvert, 2000; Metternicht, 2001; Adriaenssens et al., 2006; Van Broekhoven et al., 2006).

Fuzzy systems use linguistic descriptions such as ‘low’, ‘high’ or ‘moderate’ for the quantification of variables and use ecological expert knowledge to transform these descriptions into a mathematical framework in which suitable data processing can be performed (Kampichler et al., 2000). This turns fuzzy systems into a popular technique for ecological modelling, resulting in numerous applications (Adriaenssens et al., 2004). This was illustrated by Adriaenssens (2004) in a previous study, in which an expert knowledge base was developed, structuring preferences as well as tolerances of 86 macroinvertebrate taxa for a limited set of environmental variables. These variables were selected based on expert knowledge and on multivariate analyses. The expert knowledge linked the environmental variables with the habitat preferences of the different species and was derived from eight information sources described by Adriaenssens (2004). The features of these information and knowledge resources differed in the used identification level, the source of this information, the geographical region for which this information could be applied, the groups of macroinvertebrates covered and the type of coding applied to express the environmental preferences. For each invertebrate species, the expert knowledge base was obtained by averaging the habitat preferences of this species as described in the eight different information sources. Although validation of these expert models on Zwalm river basin data and on data from the Overrijsel Province (the Netherlands) showed that many models adequately described the invertebrate habitat preferences, a substantial number of models poorly reflected the invertebrate habitat preferences. Therefore, Van Broekhoven (2007) applied a data-driven genetic algorithm to optimise the fuzzy sets of these models. The results showed that most of the optimised models were less relevant because of the loss of interpretability after the optimisation. Moreover, this approach assumed that the expert knowledge, and thus the expert rule base, is consistent over the different ecological situations described in the Overrijsel Province dataset.

However, several authors contest the consistency of this expert knowledge over different ecological situations (Gore and Nestler, 1988; Vismara et al., 2001; Acreman and Dunbar, 2004; Vilizzi et al., 2004; Moir et al., 2005; Vilizzi and Copp, 2005). Moreover, the formalisation of problem-relevant human expert knowledge is often difficult and tedious, and thus the main bottleneck in the application of fuzzy logic is the need for ecological expert knowledge. Recent research has shown that complementing fuzzy systems by data-driven techniques can solve this knowledge acquisition bottleneck (Chen and Mynett, 2003; Žnidaršic et al., 2006). For example, the induction of fuzzy rule-based models by heuristic search algorithms is often used in the field of fuzzy rule learning (Hüllermeier, 2005).

This dissertation aims to develop and test a data-driven optimisation method for fuzzy habitat suitability models. Specifically, a hill-climbing algorithm was applied to optimise the fuzzy expert rules of the model and the impact of different training criteria on the optimisation results was analysed. The fuzzy sets were optimised based on their entropy, which quantifies the uniformity of the data distribution over the input space. The value of the presented approach for integrated river management was tested on 5 datasets which described ecological case studies in Belgium, Switzerland and New-Zealand. Although the term ‘species distribution model’ may indicate a wider context than the term ‘habitat suitability model’, both terms are used as synonyms in this dissertation (see also Section 2.2 for further discussion).

In the first case study, a fuzzy habitat suitability model for macroinvertebrates in rivers is developed at the mesohabitat scale level. The ecological expert knowledge needed for this model is derived in two different ways in order to **compare expert knowledge derived from literature to knowledge derived from data** using a hill-climbing algorithm. Habitat models were generated for the mayfly *Baetis rhodani* Pictet in the Zwalm river basin (Belgium), which is an indicator species for a good ecological quality. This study aims to demonstrate the importance of ecological data in the model development process.

To develop reliable data-driven models, however, a sound comparison of model performance is needed to allow reliable model training and evaluation of the final model. A crucial step in the model comparison procedure is the assessment of the model performance (Fielding and Bell, 1997; Manel et al., 1999a; Manel et al., 1999b; McPherson and Jetz, 2007). The key component of model performance assessment is the performance criterion applied to quantify model performance. In the second part of this dissertation, the **role of these performance criteria in the evaluation process of the final model** is reviewed and some empirical insights are given which are important for both model training and evaluation. Furthermore, the role of performance criteria **in model training** is analysed theoretically and empirically in the second case study. Specifically, a habitat suitability model was developed for spawning grayling in the Aare river (Bern, Switzerland) and the effect was analysed of two different performance measures, the weighted % Correctly Classified Instances (% CCI_w) and Cohen's *Kappa* (Cohen, 1960), on the training results.

Although the CCI_w allows for **prevalence-adjusted model training**, this criterion does not take into account the specific characteristics of a fuzzy species distribution model. Therefore, this dissertation presents the adjusted average deviation (*aAD*), which is similar to the average deviation (*AD*; Van Broekhoven et al., 2007) but contains a parameter α to account for different prevalences of the model training set. Previous research showed that the *AD* is an appropriate performance measure for fuzzy models because it takes into account the fuzzy characteristics of the model output (Van Broekhoven et al., 2007). Specifically, the *AD* takes into account the order of the output classes by returning the average deviation between the position of the class obtained with the model and the position of the class stored in the data set. In contrast to the *aAD* however, the *AD* does not distinguish omission from commission errors and does not allow prevalence-adjusted model training.

The *aAD* is applied on the second and the third case study to analyse the strengths and weaknesses of the *aAD* for ecological modelling. The relation between the parameter α and the training set prevalence was analysed for grayling in the Aare River and for three other species in different New Zealand river systems: the caddisflies *Aoteapsyche* spp., large brown trout *Salmo trutta* and rainbow trout *Oncorhynchus mykiss*. Therefore, the *aAD* was implemented in a hill-climbing algorithm to optimise a fuzzy species distribution model for each species.

The focus of this dissertation is on presence-absence models, but some concepts and problems which are being discussed also apply to ecological modelling more generally. Moreover, recommendations are given on the application of different performance criteria on model evaluation and training. Due to the generality of the developed models, the presented methods could be easily applied to other species, communities or ecosystems. Therefore, this dissertation may contribute to the development of more reliable species distribution models and to the implementation of these models into integrated river management.

1.2. A road map to this dissertation

This dissertation consists of three main parts as shown in Fig. 1.1. The first part includes introductions to species distribution models, fuzzy rule-based models and data-driven optimisation of these models. The second part considers the difference between expert knowledge-based and data-driven model development and includes the first case study. The third part focuses on prevalence-adjusted model training for data-driven model development and contains the second and the third case study.

The first part consists of three chapters. Chapter 2 starts with an introduction to species distribution modelling, which describes the importance and the state-of-the-art of species distribution models. Furthermore, the challenges for these models are discussed and the knowledge gaps which are significant for this dissertation are highlighted. Specifically, Chapter 2 illustrates how fuzzy models could contribute to species distribution modelling. In Chapter 3, these fuzzy rule-based models are described into detail. This chapter provides the reader with an introduction to fuzzy set theory, which is followed by a description of the fuzzy rule-based model type which is applied in this dissertation: the linguistic fuzzy models, including the Mamdani-Assilian models. This chapter aims to familiarise fuzzy modellers with the notations used in this dissertation and to provide other readers with a sufficient stock-in-trade concerning fuzzy modelling. Chapter 4 describes the fuzzy rule-based optimisation techniques which were developed and applied in this dissertation. This chapter can be split into three main sections: the first section describes the fuzzy set optimisation method which was developed in this dissertation, while the second one discusses the data-driven hill-climbing algorithms for rule base optimisation. This section also presents the performance criteria which are being applied most frequently to optimise and evaluate species distribution models. The last section considers the input variable selection approach which was developed in this dissertation. In contrast with the first two sections which are applied on all further case studies, the third one is only applied on the first case study, which is described in Chapter 5. Therefore, the last section of Chapter 4 is included in the second part of this dissertation, which analyses the difference between expert knowledge-based and data-driven fuzzy rule-based models. This is illustrated in Chapter 5, in which a fuzzy habitat suitability model for macroinvertebrates in rivers is developed. The ecological expert knowledge needed for this model is derived in two different ways in order to compare expert knowledge derived from literature with knowledge derived from data using a hill-climbing algorithm. In this chapter, species distribution models are generated for the mayfly *Baetis rhodani* in the Zwalm river basin (Belgium), which is an indicator species for a good ecological water quality. The hypothesis tested hereby is that data-driven models outperform expert knowledge-based models. In this chapter, abundance data are being applied, whereas presence-absence data are being applied in all other chapters. The reason for this is twofold: first, the ecological expert knowledge used in Chapter 5 was developed for abundance data, and second, the modelling approach developed in this presentation can deal with any output variable format (abundance, presence/absence, ...) due to its flexibility and generality.

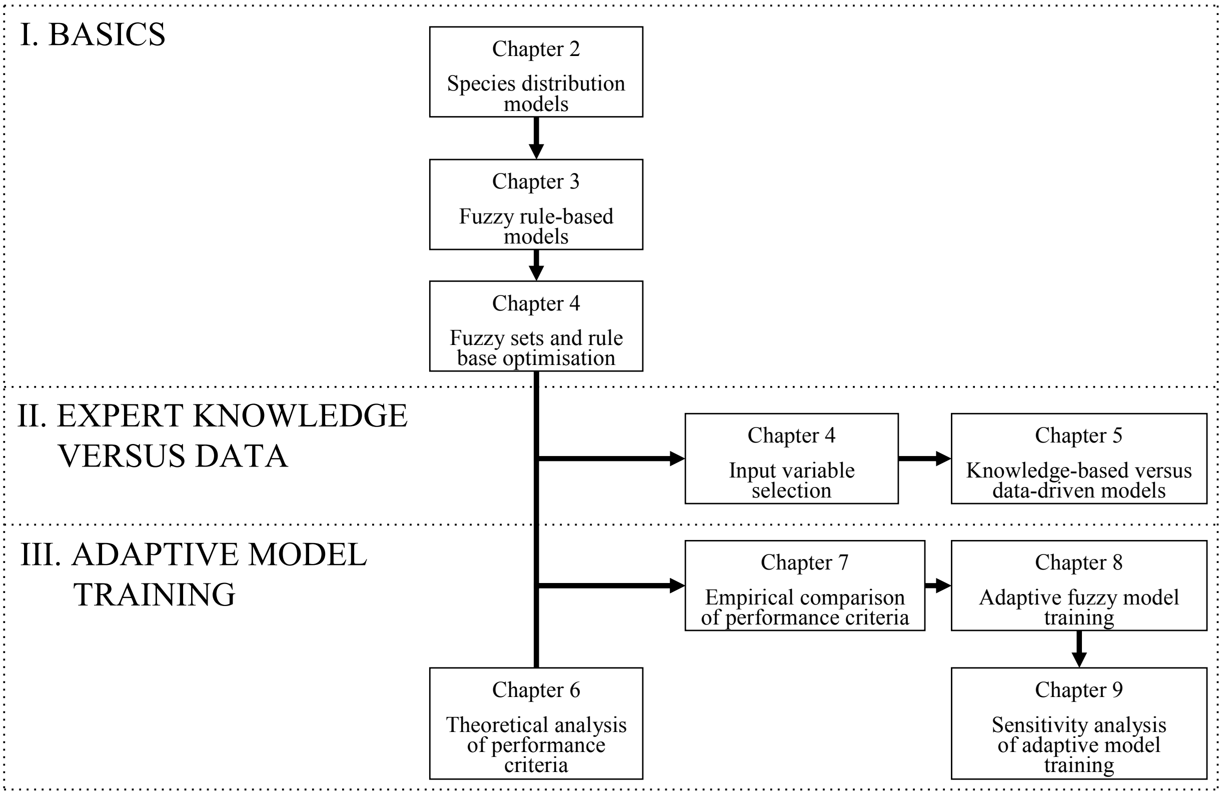


Fig. 1.1. A road map to this dissertation.

The third part of this dissertation focuses on the criteria which quantify the performance of presence-absence species distribution models. Chapter 6 provides a conceptual description of model performance and discusses the state-of-the-art on the performance assessment of species distribution models. Furthermore, this chapter theoretically analyses the role of performance criteria in the model training and evaluation process. The aim of this chapter is to provide the reader with a general introduction to model performance assessment, which is essential to comprehend Chapters 7 – 9. The theoretical analysis of Chapter 6 is illustrated in Chapter 7 with an empirical analysis of the impact of two of the most frequently applied performance criteria, *CCI* and *Kappa*, on the model training results. This analysis is performed on the second case study, which describes a species distribution model for spawning European grayling (*Thymallus thymallus*) in the Aare river (Switzerland). Chapters 8 and 9 focus on a new concept in species distribution modelling: prevalence-adjusted model training or optimisation. Since the most commonly applied performance criteria do not take into account the fuzzy characteristics of the fuzzy species distribution models, Chapter 8 introduces the adjusted average deviation (*aAD*). This performance criterion takes the order of the fuzzy output classes into account and also allows for prevalence-adjusted model training because a parameter α is implemented in the criterion. The relation between this parameter and the prevalence of the training data set is illustrated in the second (Chapter 8) and the third case study (Chapter 9). In the latter study, species distribution models are being developed for three species in different New Zealand river systems: the caddisflies *Aoteapsyche* spp., large brown trout *Salmo trutta* and rainbow trout *Oncorhynchus mykiss*. *The hypothesis tested hereby was that there exists a universal relation between the value of* α and the prevalence of the training data set. The dissertation concludes with general conclusions and suggestions for future research in Chapter 10.

CHAPTER 2

Species distribution models

2.1 Importance of species distribution models

Although many studies have assessed the environmental responses of river communities or taxa to specific conditions, there is a clear need for models that quantify the species-environment relationships (Guisan and Zimmermann, 2000). Species distribution models, habitat suitability models or ecological niche modelling allow for such quantification because these models predict species distributions based on data describing the abiotic environment. River management and conservation biology can benefit from these predictive models as decision support tools for adaptive management (Guisan and Zimmermann, 2000; Austin, 2002; Scott et al., 2002). Specifically, habitat suitability models may select the most optimal river restoration scenario based on quantitative prediction of river restoration effects. These models can also play an important role in basic ecological research, for example by quantifying the river status, assessing the anthropogenic impacts or revealing ecological networks and knowledge. As such, predictive ecology stimulates the elaboration of scientific knowledge, which could lead to better understanding and optimised river monitoring and management. As an example of the contribution of habitat models to monitoring, (Guisan et al., 2006) apply these models to improve the sampling of rare species. In past decades, habitat suitability models have increasingly received attention due to their wide management applications in the context of biogeography, conservation biology and climate change studies (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005; Araújo and Rahbek, 2006; Dominguez-Dominguez et al., 2006). Consequently, habitat suitability models can contribute substantially to society by keeping the cost-efficiency ratio for river management as low as possible.

2.2 State of the art

Until 1980, limnology was mainly descriptive and diagnostic, with a limited interest for prediction. Subsequently, various research fields arose, which can now be joined under the field of ecological modelling: habitat suitability modelling, bio-energetic modelling, population dynamics modelling, process modelling, spatial and landscape modelling, ecotoxicological modelling and ecological-economical-social modelling (Jørgensen, 2005). Although all these models are inherently related and should be combined to create a truly predictive ecological model, this dissertation focuses on the field of habitat suitability modelling. Habitat suitability models predict species distributions based on data describing the abiotic environment. As such, the terms “habitat suitability model” and “species distribution model” are treated as synonyms in this dissertation because a habitat is assumed to be suitable if a species is observed at this habitat, while a species is assumed to be absent at an unsuitable habitat. However, species absence at a specific habitat does not imply that the habitat is unsuitable; neither does it indicate that a species will never occur at this habitat. This issue is further discussed in Chapter 6.

Depending on the aquatic subecosystem which is modelled, different types of habitat models can be distinguished. Physical habitat models link the physical characteristics of rivers to the habitat suitability of an aquatic species of interest. Other models provide a more integrated approach since they include chemical water quality or trophic relations between species. Finally, fully integrated models consider wider aspects such as economical, social or policy issues.

2.2.1 Two major trends in river species distribution modelling

Physical habitat suitability models are treated as a separate model category in this dissertation because the physical habitat is the key factor which distinguishes running waters from other aquatic ecosystems. Since their development in the eighties, physical habitat models have become an important tool for river management (Armour and Taylor, 1991; Bockelmann et al., 2004). The models allow evaluating habitat suitability for aquatic organisms, based on physical variables such as depth, flow velocity and substrate (Bovee, 1982). The influence of flow changes on biological diversity can be analysed since these variables are depending on the water regime (Holm et al., 2001). This is particularly useful for impact assessment of hydropower plants or water abstraction on the ecological river status and for determining the minimum flow requirements of aquatic populations. Physical habitat models can also be used to simulate and evaluate the effect of restoration projects by adjusting the state of the input variables (Shuler and Nehring, 1994; Shields et al., 1997).

Since the late seventies, aquatic habitat simulation models have been used for fish in water resource management, particularly in North America (Bovee, 1982). The Instream Flow Incremental Methodology (IFIM) (Bovee, 1982; Stalnaker, 1995) and its PHABSIM component (Bovee, 1982; Milhous et al., 1989) are considered to be the first of these fish habitat models and are now being applied worldwide. In northern Europe for instance, the PHABSIM approach has been used for the evaluation of restoration actions in the nineties (Huusko and Yrjana, 1997). Other models that were based on PHABSIM include the Norwegian River System Simulator (Alfredsen and Killingtveit, 1996), RHYHABSIM (Jowett, 1996), and EVHA (Ginot, 1995). (Parasiewicz, 2001) developed the Mesohabitat Simulation Model (MesoHABSIM) to address some shortcomings of PHABSIM related to its application on larger scales (Parasiewicz, 2007b; Parasiewicz, 2007a; Parasiewicz and Walker, 2007). Newer applications of PHABSIM are also including 2D hydrodynamics (Katopodis, 2003).

The PHABSIM approach was originally developed for the purpose of instream studies with a focus on river hydraulics and initially only considered physical and structural variables as descriptors of changes in species, at the level of density or biomass (Stalnaker, 1995). PHABSIM relates open channel hydraulics with known elements of the fish and macroinvertebrate communities. All these models link physical variables such as flow velocity, depth or bottom substrate composition to habitat suitability by means of uni- or multivariate preference functions (Bovee, 1982). From these descriptors it is possible to estimate the 'Weighted Usable Area' (WUA) or the 'Weigthed Usable Volume' (WUV; (Mouton et al., 2007a) of a species in a stretch as a function of flow (Gore and Nestler, 1988). These habitat preference functions have two major disadvantages: many possible interactions between physical habitat variables are ignored, while the spatial links between habitats are neglected (Jorde et al., 2000). Another shortcoming is the fact that macroinvertebrate communities in many sites are more affected by water quality than by physical habitat conditions (Karr, 1991).

The Computer Aided Simulation Model for Instream Flow Requirements (CASiMiR; (Jorde et al., 2000; Schneider, 2001) uses a new approach, based on fuzzy sets and rules. These linguistic rules describe the imprecise information which is often characterising ecological data and enable implementing the interactions between physical variables. Although several fish species prefer certain types of cover in shallow areas for instance, these cover types can be neglected in deeper zones. Furthermore, the knowledge about habitat requirements of fish often consists of semi-quantitative data. The habitat selection of several species can be predicted adequately based on long-time experience of fish biologists and on analysis of monitoring data. However it is observed that habitat requirements of species depend on the life stage and river type (Jungwirth et al., 2000). Related to the preference functions of PHABSIM, expert rules can be defined in the fuzzy model for different species and life stages (Schneider, 2001).

Since the start of physical habitat modelling, several models have been developed which attempt to provide a more **integrated approach to habitat modelling**. Apart from physical habitat variables, these models consider other factors which can affect habitat suitability of aquatic species. As such, these models claim to be of greater value for river management than physical habitat models since the latter only focus on a specific aspect of the river ecosystem. However, the pitfall of these integrated approaches lies in the trade-off between complexity and accuracy. For the sake of simplicity, many integrated approaches only include simplified environmental processes. For example, they include the oxygen concentration measured at one moment in time to represent the availability of dissolved oxygen to the aquatic species. Such simplifications can be useful but should be treated with care and should not be generalised before sufficient knowledge of the environmental processes is available to test the model results. Specifically, a modular habitat suitability model in which each complex environmental process is described by a specific module, seems a more promising and more reliable approach to support integrated river management.

The River Invertebrate Prediction and Classification System (RIVPACS; Wright et al., 1984) was one of the first integrated models with a predictive capacity for use in river management. It was developed for assessing the ecological quality of rivers in the UK and was based on the presence of macroinvertebrate communities (Wright et al., 1984). RIVPACS develops statistical relationships between the fauna and the environmental characteristics of a large set of high quality reference sites. These relationships can be used to predict the macroinvertebrate fauna which is expected to occur at any site in the absence of pollution or other environmental stresses. The observed fauna at new test sites can then be compared with their site-specific expected fauna to derive indices of ecological quality (Clarke et al., 2003). Later on, AUSRIVAS (Australian River Assessment System; (Norris and Norris, 1995) and

BEAST (Benthic Assessment of Sediment; (Reynoldson et al., 1995) were developed in Australia and in North America respectively based on the RIVPACS approach. However, RIVPACS cannot be used as a dynamic model to predict the impact of environmental changes because data from impaired sites are not included (Clarke et al., 2003).

Although many detailed studies allow for the prediction of the effectiveness of different restoration options for improving ecological river quality (Jowett, 1997; Parasiewicz and Dunbar, 2001; Roni et al., 2002; Palmer et al., 2005; Mouton et al., 2007b), the fine scales at which these studies are conducted need to be reconciled with the coarse scale at which river managers make decisions (Petts et al., 2006). Hence, apart from being integrated, habitat models should be communicative tools which allow bridging the gap between the results of detailed studies on water systems and the information currently used in water management (Roux et al., 2006). Specifically, such tools should implement previous research findings and allow for the comparison of different restoration scenarios to support decisions of river managers (King et al., 2003).

The communicative aspect of such tools is one of the key factors for their success. Not only should they enable river managers and scientists to justify their decisions to other stakeholders, but they should also enhance communication between these groups to extend their general knowledge of the river system. Rogers (2006) emphasised that the development of this collective understanding and integration within and between scientists, river managers and stakeholders is the real challenge in river management. This ‘process of relating’ (Roux et al., 2006) should stimulate all partners to learn and avoid prejudices or misunderstandings between managers and scientists (Ludwig, 2001).

The Water Framework Directive-Explorer (WFD-Explorer) is a user-friendly tool which was developed for this purpose and successfully implemented in river management in The Netherlands (Van der Most et al., 2006; Mouton et al., 2008). The toolbox implements ecological knowledge using rules which link variables such as flow velocity and nutrient concentrations to the ecological quality of the different water bodies in the river basin of interest. The current river status and the effects of different restoration scenarios are analysed at the river basin level, thus supporting river managers in the development of the river basin management plans needed for WFD implementation (Mouton et al., 2008).

2.2.2 Conceptual aspects of species distribution modelling

General features of species distribution models

The various steps of the development process of predictive modelling were described by several authors (Jørgensen, 1999; Manel et al., 1999; Guisan and Zimmermann, 2000; Jørgensen, 2000; Olden and Jackson, 2002; Robertson et al., 2003; Jørgensen, 2005). The features that are important for the models applied in this dissertation, will be discussed in the following paragraphs.

Apart from the two aforementioned major trends in habitat suitability modelling, habitat models can also be categorised based on their conceptual basis. Guisan and Zimmerman (2000) posed that all habitat models are based on more or less the same **framework**, which commonly consists of three major components (Austin, 2002). The first one is an ecological model, which includes the ecological knowledge and theory to be used and tested in the study, while the second one is a data model, describing the decisions to be made regarding how the data are collected and how the data will be monitored. The third component is referred to by (Guisan and Zimmermann, 2000) as the statistical model and includes the error function and significance testing. However, with the rise of new powerful computational techniques and GIS tools, the term ‘computational model’ would be more accurate for the third component.

Many habitat models relate the geographical distribution of species or communities to their actual environment, hereby assuming that the modelled communities are in equilibrium with their environment (Guisan and Theurillat, 2000; Guisan and Zimmermann, 2000; Pulliam, 2000; Tyre et al., 2001; Barry and Elith, 2006). Consequently, most habitat suitability models are **static**, while the accuracy of prediction will depend on the degree to which dynamic processes are important in the ecosystem of interest (Austin, 2002). Although oligotrophic stream systems are harsh environments containing fewer competitive dominants and therefore having less needs for interspecific resource partitioning (Townsend, 1989; Hildrew, 1992), biotic interactions in general need to receive more attention (Guisan and Zimmermann, 2000). The non-equilibrium concept is thus more realistic in ecology (Pickett et al., 1994; Austin, 2002; Austin, 2007), because it includes equilibrium as a possible state. Yet, dynamic simulation models (Korzukhin et al., 1996) require intensive knowledge of the species involved, whereas static models allow large-scale prediction with less effort, and the advantage that no detailed knowledge of the physiology and the behaviour of the species is necessary (Guisan and Zimmermann, 2000).

Depending on the overall goal of the modelling effort, habitat suitability models can focus on three desirable model features: **generality**, **precision** and **reality** (Levins, 1966). General models are robust across different environments or scenarios, while precise models can accurately indicate the habitat suitability for a species. The precision of a model is related to model uncertainties *sensu lato*: these uncertainties can be associated with the model output (Omlin and Reichert, 1999), but also with other aspects of the modelling process (see Brugnach et al. (2007) and Pahl-Wostl et al. (in press) for a full discussion). The reality of a model indicates to which extent natural processes are represented in the model. Many habitat suitability models sacrifice generality for high reality and precision, although several authors argue that precision, generality and reality are not always mutually exclusive (Guisan and Zimmermann, 2000).

Data collection and variable selection

Austin (1985) and Austin et al. (1984) defined three types of ecological gradients, namely resource, direct and indirect gradients. Resource gradients consist of matter and energy which is consumed by species (e.g. nutrients, water, light), while direct gradients are environmental variables that have physiological importance, but are not consumed, such as temperature, water column depth, flow velocity or bottom substrate composition. Whereas an direct variable is the most proximal gradient in the chain of processes that link the variable to its impact, indirect variables are clearly distant variables, but are linked more closely to the causal variable with regard to human impact (Austin, 2007). Different resources and direct gradients usually can be replaced in a simple way by one indirect variable (river bank condition, river sinuosity, slope, stream order) (Guisan et al., 1999). Although the use of direct and resource variables ensures that the model is more general and applicable over larger areas, these variables are usually more difficult or more expensive to measure than indirect variables (Guisan and Zimmermann, 2000).

The quality of the habitat model and the **sampling design** influence each other significantly. Consequently, the formulation of the conceptual model can be helpful to design an efficient sampling strategy. Several authors argue that, if a quantitative analysis between species distribution and environmental variables is the focus, a stratifying sampling approach is most useful (Guisan and Zimmermann, 2000; Hirzel and Guisan, 2002). Specifically, one will attempt to sample an equal number of replicates per environmental combination. Hirzel and Guisan (2002) also stress the effect of sample size on the model performance, especially for statistical models due to the relation between sample size and the number of degrees of freedom. Moreover, a minimal distance between two sampling sites should be defined prior to sampling to avoid spatial autocorrelation (Legendre, 1993) and a further reduction of the number of degrees of freedom.

The conceptual model may also lead to the choice of an appropriate **spatial scale** for conducting the study (Maddock, 1999; Guisan and Zimmermann, 2000). For running waters, different scale levels can be distinguished: the river basin or macrohabitat, the stream, the river stretch or mesohabitat, and the microhabitat (Maddock, 1999). Several authors have proposed the mesohabitat as an appropriate scale level for assessing the bottlenecks in a river system (Maddock, 1999; Parasiewicz, 2007a; Parasiewicz and Walker, 2007) because it provides an intermediate approach between the coarse river basin scale and the detailed but more time consuming micro scale habitat assessment. Specifically, a mesohabitat is defined as a species- and life-stage specific area where the configuration of hydraulic patterns together with attributes that provide shelter create favourable survival and development conditions (Parasiewicz, 2007b). For many fish species for instance, the shallow fast-flowing river patches are referred to as riffle mesohabitats, whereas the deeper, almost stagnant areas are specified as pools. Different models operating at the meso scale have been developed such as MesoHABSIM (Parasiewicz, 2007b) and MesoCASiMiR (Mouton et al., 2006) which is based on fuzzy logic (Zadeh, 1965). Finally, the model performance may also depend on **temporal scale** effects, such as the choice of a sampling period, duration or frequency (Mouton et al., submitted). Several physical habitat models use data sampled at different flows to allow prediction of the habitat suitability in the whole flow range between base flow and peak flow (Parasiewicz and Walker, 2007). Another temporal scale effect is seasonality, which may be an important aspect when aquatic species are sampled (MakiPetays et al., 1997; Nykänen et al., 2001; Nykänen et al., 2004).

2.2.3 Habitat suitabilty

General definition of habitat suitability

This dissertation focuses on habitat suitability as the biological response of aquatic species to environmental conditions because habitat suitability is the key factor explaining species distribution. Nevertheless, the biological response can be expressed by many different other measures such as species presence, abundance, density, usable area or volume. The use of these measures in habitat modelling depends on the model objective and the constitution of the available data. Although species presence or absence is the most easily assessable, more biological information can be derived from species abundance or density. The application of species abundance originates from the niche theory (Hutchinson, 1957) and has been proved a good expression of habitat suitability for macroinvertebrates (Statzner et al., 1988; Gore et al., 2001) and fish (Bovee, 1982). Species density is a less common measure in habitat suitability due to its hard and time-consuming assessment in the field and its limited added value for river management.

Species response curves or habitat preference curves show the response of a species to an environmental gradient and thus the shape of these curves will vary with the nature of the gradient. Although the response to a direct gradient is assumed to be fixed and bell-shaped (Hutchinson, 1957), the response to an indirect gradient could take any shape because this shape depends on the correlations between the indirect variable and the causal gradient (Guisan et al., 1999; Guisan and Zimmermann, 2000; Meentemeyer et al., 2001). Moreover, whether the direct effect of a variable on the species response in a specific river stretch can be observed is situation-dependent. For instance, the effect of the direct gradient ‘flow velocity’ on caddis fly habitat suitability could be clearly observed in rivers with a good chemical water quality, whereas the effect of this variable will be far less obvious in polluted rivers. The species response at each value of the environmental gradient is referred to as the habitat suitability index (HSI) for this environmental condition (Gore and Nestler, 1988). From these HSIs, the Weighted Usable Area (WUA) or the Weighted Usable Volume (WUV) of a species can be calculated, as described in Box 1. In literature, ‘species response’, ‘habitat preference’ or ‘habitat suitability index’ are often used as synonyms and consequently these terms will be referred to in this dissertation as habitat suitability. In addition to univariate species response curves, *n*-dimensional response surfaces can be constructed for combinations of variables. Many physical habitat models use two-dimensional surfaces for combinations of the most common variables such as depth, flow velocity and bottom substrate (Bovee, 1982; Gore and Nestler, 1988; Parasiewicz, 2007b).

The Physical Habitat Simulation approach

The most common physical habitat models, such as PHABSIM (Physical HABitat SIMulation; Bovee, 1982; Elliott et al., 1999), combine the Usable (habitat) Area (Osborne and Suarez-Seoane) with these HSIs in order to define an area of habitat suitable for the species of interest, the Weighted Usable Area (WUA) (Waters, 1976). The UA is hereby based on two-dimensional (2D) distributions of habitat features in the horizontal plane of the river surface or river bed. In practice, the river bed is divided into cells and each cell is characterized by specific values of the physical variables. Based on these values, an HSI is attributed to each cell. The WUA, as applied in PHABSIM model studies, is then obtained by integrating the habitat quality over the cells of the entire studied stretch:

$$WUA = \sum_{i=1}^{n} A_i \cdot HSI_i \ = f(Q) \qquad (m\ ) \qquad (2.1)$$

with A_i the horizontal surface of cell i, HSI_i the habitat suitability index of cell i and n the number of cells in the studied river stretch. A great deal of research has been done on the application and adaptation of these physical habitat modelling approaches (Capra et al., 1995; Heggenes, 1996; Parasiewicz and Dunbar, 2001; Mouton et al., 2007a; Parasiewicz, 2007b).

Several aspects of these approaches have been criticised since the eighties and numerous specific modelling applications have demonstrated some improvement (Acreman and Dunbar, 2004). Many authors suggested that microhabitat use of stream fish and invertebrates, and hence the HSIs of these species, vary at different spatial and temporal scales (Gore and Nestler, 1988; Vismara et al., 2001; Acreman and Dunbar, 2004; Vilizzi et al., 2004; Moir et al., 2005; Vilizzi and Copp, 2005). Other critique considers the representation of the three-dimensional flow environment (Mouton et al., 2007a). Since flow velocity changes with depth and consists of different values for each point in the river, all these pseudo 2D approaches use the depth averaged flow velocity as an estimation of flow velocity at a single point. However, the use of mean velocities to define the fluvial habitat can be misleading as two points may have similar depth averaged flow velocities, but sharply contrasting velocity profiles (Beebe, 1996; Stalnaker et al., 1996; Mouton et al., 2007a).

Increasingly, the three-dimensionality of habitat hydraulics is being recognised as an essential issue for understanding the ecological needs of aquatic species (Ghanem et al., 1996; Bremset and Berg, 1999; Newson and Newson, 2000; Rhoads et al., 2003). Several authors have suggested that particular water depths are preferred habitats for specific aquatic species or life stages (Greenberg et al., 1996; Heggenes, 1996; Bremset and Berg, 1999). Others indicate the importance of flow velocity at different depths (Kemp et al., 2003). New methods to quantify physical habitat features have been published (Nestler and Sutton, 2000), while greater hydraulic process representation may be achieved using 2D and 3D hydrodynamic models (Booker, 2003; Pasternack et al., 2004; Stewart et al., 2005). These models quantify flow velocities at different depths, allowing linkage with bioenergetic models (Rosenfeld, 2003; Booker et al., 2004). Consequently, Mouton et al. (2007a) have defined the physical habitat as a combination of different habitat features based on a three-dimensional characterisation of the flow environment. They describe the formulation and testing of a physical habitat model which, instead of the UA, calculates the Usable Volume (UV), which is the volume of usable habitat for a species of interest (Mouton et al., 2007a). The resulting model generated the Weighted Usable Volume of aquatic species and was calibrated and validated in an artificial river and in two natural river stretches in Austria and Belgium (Mouton et al., 2007a).

Conceptual remarks on habitat suitability

In contrast to the habitat suitability of individual species which is modelled in this dissertation, other research focuses on community descriptors such as environmental indices (Rosenberg and Resh, 1993; Mouton et al., 2008), biological criteria (Kearns et al., 1992) or species richness (Mac Nally et al., 2003). However, modelling species instead of communities comes closer to what is believed to meet reality (Gleason, 1926; Guisan and Zimmermann, 2000). A general consensus has arisen that accepts the continuum concept (Gleason, 1926) in preference to the community concept (Clements, 1916). Specifically, the continuum concept states that most communities change continuously along environmental gradients, rather than forming distinct, clearly separated zones (Whittaker, 1975). This continuous change is explained by the individualistic hypothesis of Gleason (1926): each species is distributed in its own way because it is affected by its interactions with the physical environment, by population-level interactions between species and by disturbance regimes. Ecological succession is thus no longer seen as having a stable end-stage called the climax, which has also its implications for the definition of reference sites (Austin, 1985). As an extension of the species approach, the species traits approach (Resh et al., 1994; Usseglio-Polatera, 1994) investigates the relation between the abiotic environment and species traits such as size, mobility or body form (Townsend et al., 1997). Based on these species traits, non-taxonomic aggregations of taxa into behavioural, life historical and functional categories such as functional feeding groups (Cummins, 1973) have been used in stream ecological studies.

The species traits approach is related to the choice of a species-based or genus/family/group-based prediction. This issue is particularly important for invertebrate species because their identification to the species level is much harder than it is for fish. Since the taxonomy of aquatic species is often based on morphological features and not on functional characteristics, species within a genus may have different ecological requirements and the response of species within a genus to different impacts may differ significantly (Schmidt-Kloiber and Nijboer, 2004). Although some studies that use invertebrates as water quality indicators obtain similar results with family or finer (species and genus) level of identification (Wright et al., 1993), the concept of the species as the basic biological unit is widely accepted (Schmidt-Kloiber and Nijboer, 2004).

Focusing on the species level, it is important to differentiate between the fundamental and the realized niche of an aquatic species (Guisan and Zimmermann, 2000). A niche of an organism is the status of an organism within its environment and community, which affects its survival as a species. Specifically, the habitat of a species indicates the abiotic conditions in which a species occurs, whereas the niche describes how a species interacts with both abiotic and biotic conditions (Hutchinson, 1957). The fundamental niche is primarily a function of physiological performance and ecosystem constraints, while the realized niche additionally includes interactions and competitive exclusion (Malanson, 1997). Consequently, the difference between both niches distinguishes whether a simulated distribution is predicted from theoretical physiological constraints or rather from field-derived observations (Guisan and Zimmermann, 2000). The fundamental niche is also referred to as the habitat preference or the habitat suitability, whereas the realized niche is equal to the habitat selection or use. There has been much debate on the assumption that the fundamental niche of a species is constant, which is often referred to as ‘niche conservationism’, and was reviewed in Wiens and Graham (2005). The models in this dissertation can describe both niches because they can deal with either expert knowledge, observed data, or a combination of both.

2.2.4 Common modelling techniques in species distribution models

In addition to the aforementioned PHABSIM technique, several other modelling techniques have been applied in habitat suitability modelling. These techniques can be subdivided into three main categories: data-driven models, expert knowledge approaches, or a combination of both, referred to as hybrid models.

Data-driven models derive the link between abiotic characteristics and habitat suitability from data which are collected in the stream of interest. These models apply common data mining techniques such as artificial neural networks, decision trees or support vector machines which can be either black-box or grey-box techniques. Black-box approaches, such as neural networks and support vector machines, deal with structures and associated parameters that usually do not have any physical significance (Babuška and Verbruggen, 1997). Artificial neural networks, for instance, are created by optimising weights and bias parameters which do not represent any physical variable. Although some authors developed networks which are more easily interpretable (Jaarsma et al., 2007), the quality of most neural networks can only be evaluated by cross-validation and comparison of model inputs and outputs (Fielding and Bell, 1997). The more interpretable approach is the white box modelling, which also allows evaluation of the model structure because this structure is physically significant. White-box models assume that the system’s behaviour is fully known, and that there exists a suitable mathematical scheme, for instance a set of differential equations, to represent this behaviour. Practical limitations of the white-box approach arise when the underlying phenomena are poorly understood, the values of various system parameters are inaccurate, or when the resulting model is highly complex (Casillas et al., 2003). Consequently, grey-box models create an intermediary third approach which attempts to combine the advantages of the white-box and the black-box approaches. Specifically, the known parts of the system are being modelled using a priori knowledge about the system, whereas the less known parts are described with black-box procedures. Decision tree models, for example, may not only generate model output, but also a highly interpretable decision tree from which the model output is derived.

The second category of habitat suitability models is based on expert knowledge. This knowledge is derived from literature or from experts and stakeholders, and describes the relation between the abiotic environment and the habitat suitability. Common techniques applied in this model category are fuzzy logic and Bayesian belief networks, but also decision trees could be derived from literature. Expert knowledge-based models, or simply expert models or knowledge-based models, are meaningful to the broad range of persons involved in the modelling process because these models allow a clear presentation of the model structure and the inference process (Borsuk et al., 2004). Hence, the model interpretability is one of the main features of knowledge-based systems, which can be an excellent basis for collaboration between river managers, scientists and other stakeholders (Casillas et al., 2003; Adriaenssens et al., 2004). Moreover, neglect of ecological knowledge is still a limiting factor in the application of statistical modelling in ecology and conservation planning (Austin, 2002; Borsuk et al., 2004).

Recent research has shown that complementing expert systems by data-driven techniques can alleviate the disadvantages of both (Chen and Mynett, 2003; Žnidaršic et al., 2006). Hybrid models are thus not only easily interpretable, but also allow the incorporation of available data or expert knowledge, which facilitates model development. For example, the induction of fuzzy rule-based models by heuristic search algorithms is often used in the field of fuzzy rule learning (Hüllermeier, 2005). This dissertation aims to develop and apply a fuzzy optimisation algorithm as a hybrid modelling technique for habitat suitability modelling.

2.3 Challenges for species distribution models

2.3.1 Knowledge-based versus data-driven models

Although expert knowledge and data-driven methods have been described as two separate approaches by many authors, there actually is no sharp distinction between both because all expert knowledge has been derived from observations or measurements. Knowledge derived from data is often quantitative, while most sensory-perceived knowledge is qualitative. Furthermore, expert knowledge can describe environmental processes, but also the occurrence of species, which integrates several environmental processes. The subject and origin of the expert knowledge determine the niche which could be described by the knowledge (Table 2.1).

Table 2.1. Relation between subject and origin of the expert knowledge and the output of the expert model.

Subject of expert knowledge	Origin of expert knowledge	Output of expert model	
		Fundamental niche / Habitat suitability	Realized niche / habitat use
Environmental processes	Sensory perceptions	X	
	Measured data	X	
Species occurrence	Sensory field perceptions	X	X
	Measured field data	X	X

Most aquatic expert knowledge which focuses on species occurrence is derived from sensory field perceptions and thus qualitative. An example would be a fisherman saying: ‘At a high flow velocity and a low depth you’ll find many brown trout in this river’. Expert models translate this quantitative knowledge into linguistic rules (Zadeh, 1965; Van Broekhoven et al., 2006) such as ‘if flow velocity is high and depth is low then the habitat suitability for brown trout is high’.

- Expert knowledge-based models may take into account most of the ecological knowledge which is available.

2.3.2 Ecological boundaries

To construct the linguistic expert rules, the range of the model variables depth, flow velocity and habitat suitability is divided into classes such as low, medium and high. Most habitat suitability models apply the crisp boundary approach, which states for example that a water column depth lower than 0.5 m is low, whereas a depth exceeding 0.5 m is high. Consequently, the boundary between two consecutive classes is crisp because it is situated at a single value of the variable. However, transitions in ecology are not crisp but gradual, resulting in ecological gradients (Cadenasso et al., 2003a; Cadenasso et al., 2003b). Specifically, if the class boundary between a low and a high water column depth is set at 0.5 m, a depth of 0.49999 m will be classified as low, while a depth of 0.50001 m will be high, which does not match with ecological boundary theory (Strayer et al., 2003). Fuzzy logic has proven to be an appropriate expert model technique to deal with these ecological gradients because the boundaries between the classes of the input variables are overlapping in a fuzzy model and thus reflect these gradual transitions between classes (Van Broekhoven et al., 2006; Mouton et al., 2007b).

- Fuzzy logic models may be more appropriate for ecological modelling than classical modelling techniques because the overlap of the fuzzy classes reflects the ecological boundary concept.

2.3.3 Interdependence of variables

The fuzzy logic approach is addressing another shortcoming of the widely used PHABSIM related habitat models, which is the application of independent habitat suitability curves. These curves describe HS inadequately since in reality physical habitat variables are not independent (Heggenes, 1996). Since the first application of PHABSIM, different methods have been developed to overcome this problem. Several authors suggested that linking different habitat variables through simple mathematical operations is not adequate for the description of habitat suitability (Bain, 1995; Heggenes, 1996; Sekine et al., 1997). Although sensitivity analysis with different sets of univariate preference curves (Bovee, 1986) could partially solve this problem, Sekine et al. (1997) proposed the use of weighting factors to combine different habitat preferences based on different variables. Another option, the multivariate species response curves which were described earlier, are hardly used in practical applications due to several mathematical limitations (Bovee et al., 1998). Fuzzy habitat models, by contrast, take into account interactions between habitat variables in a multivariate habitat suitability (Adriaenssens et al., 2004; Van Broekhoven et al., 2006).

Other authors propose a more data-driven approach. Lamouroux et al. (1998) combined statistical distributions of physical habitat variable classes within river elements with statistical hydraulic models to reduce the efforts for field measurements (Lamouroux et al., 1995). Yet, the method is not transferable over different rivers. The combination of an artificial neural network and a generalized additive model for the prediction of density of roach (*Rutilus rutilus*) (Brosse and Lek, 2000) performed better than a multiple linear regression model (Gozlan et al., 1999), but many of these methods require large amounts of data or are not transferable to different river types (Lamouroux et al., 1998). These data requirements significantly restrict applicability of these approaches, even if data collection is facilitated by new techniques such as fish tagging. Fuzzy habitat models allow including expert knowledge into the calculation of habitat suitability, hence compensating situations where little fish data could be collected in the field.

- Fuzzy logic models may be more appropriate for habitat suitability modelling than some PHABSIM-based approaches since these fuzzy models include variable interdependence.
- Fuzzy logic models may be more appropriate for habitat suitability modelling than some data-driven models if field data are lacking.

2.3.4 The knowledge acquisition bottleneck

More recently, knowledge-based models have become a popular technique for ecological modelling, resulting in numerous applications (Adriaenssens et al., 2004). However, the main bottleneck in the application of a merely knowledge-driven approach is the need for ecological expert knowledge. Although ecological research currently produces a wealth of knowledge about the habitat requirements of the various species, the formalisation of problem-relevant human expert knowledge is often difficult and tedious (Adriaenssens, 2004; Žnidaršic et al., 2006).

First, most of the available knowledge is species specific, while for example habitat preference curves estimated for one species in a given stretch can often not be extrapolated to other communities of which this species forms a part (Karr, 1991). A far greater threat for the application of expert models lies in the lack of consistency of the expert knowledge, which is reported by several authors (Acreman and Dunbar, 2004; Adriaenssens, 2004; Wiens and Graham, 2005; Hernandez et al., 2006; Randin et al., 2006; Fitzpatrick et al., 2007; Strauss and Biedermann, 2007). In a study on the transferability of species distribution models for butterflies, Vanreusel et al. (2007) observed some transferability of ecological expert knowledge, but their study only covers a small study area, which minimises the likelihood of different impacts on the realised niche. Boyce et al. (2002) describe a number of biological problems that interfere with the niche conservation concept. These results are in line with the environmental gradient categorisation which was discussed earlier. Due to their direct impact on the habitat suitability of aquatic species, species response curves to direct gradients are expected to have constant shapes, whereas the shape of a species response curve to an indirect gradient may depend on the situation (Hutchinson, 1957). Consequently, expert knowledge or species response curves to a specific variable can only be consistent over two situations if this variable is a direct gradient for the species of interest in both situations. However, the same variable can be a direct variable in one river and an indirect one in another river because an direct variable was defined as the most proximal gradient in the chain of processes that link the variable to its impact. Flow velocity, for instance, has been widely accepted as a direct gradient for fish presence in rivers with a good chemical water quality, whereas flow velocity may be further down the chain of processes that link this variable to its impact in severely polluted rivers due to interactions with other variables such as oxygen concentration. Specifically, high flow velocities for instance may lead to a higher oxygen concentration due to water splashing. This oxygen concentration may thus be a direct variable for fish presence in polluted rivers, whereas flow velocity is an indirect gradient in these rivers. A general assumption of most knowledge-based habitat suitability models is that only the most direct gradients are included in the model, and the species response to these gradients is thus consistent and transferable between different rivers (Bovee, 1982). Since the distinction between direct and indirect gradients is situation-dependent, however, species response curves could take any shape depending on the studied case (Guisan et al., 1999; Guisan and Zimmermann, 2000; Meentemeyer et al., 2001; Austin, 2002; Acreman and Dunbar, 2004). This has important implications for the applicability of knowledge-based models (Hudson et al., 2003). Particular care should be taken in interpreting model results beyond their original domain (Boyce et al., 2002).

- The acquisition of expert knowledge is the key bottleneck for the application of fuzzy logic models in habitat suitability modelling.

2.3.5 Data-driven knowledge acquisition

Recent research has shown that complementing expert systems by data-driven techniques can solve this ‘knowledge acquisition bottleneck’ (Žnidaršic et al., 2006). For example, the induction of fuzzy rule-based models by heuristic search algorithms is often used in the field of fuzzy rule learning (Hüllermeier, 2005). However, ecological data show some specific characteristics which should be taken into account when developing data-driven ecological models. These characteristics are being addressed in the next chapters, with a focus on the development of reliable models and on the evaluation of these models.

The development of a reliable data-driven model requires a sound training and validation procedure. Model training is the process in which the model parameters are iteratively adjusted to increase the agreement between the model predictions and the observations, which are referred to as the training data set. Since this agreement is assessed by the performance criterion, model training aims to optimise the performance criterion. Cross-validation is often applied to assess the robustness of the training results by randomly creating different training data sets. If the number of model parameters is relatively high compared to the number of training data points, model training may result in a model which predicts the observations too accurately. Specifically, the reliability and the robustness of the resulting model decrease during model training because the model increasingly reflects the training data. The model predictions would agree poorly with the observations from other data sets because it only describes the specific ecological processes included in the original training data. In such cases, model training is a trade-off between model specificity and model generality or robustness. This situation often occurs when neural networks or decision trees are applied and is called overfitting. To avoid overfitting, the available data set is often split in a training set and a test set. Model training is then stopped when the model performance on the test set decreases (Fielding and Bell, 1997).

Although several authors emphasized the importance of correct model training and evaluation, both procedures are often neglected or applied erroneously (Fielding and Bell, 1997; Allouche et al., 2006; Goethals et al., 2007). This problem is often related to the application of the performance criterion which is used to evaluate the model performance. This criterion is the key component of model training and evaluation, and this dissertation aims to incorporate both procedures into ecological modelling in an ecologically relevant way by analysing different performance criteria and suggesting challenges for their correct application.

- Ecologically relevant model training and evaluation is a crucial step in the development process of a data-driven species distribution model.

CHAPTER 3

Fuzzy rule-based models

3.1 Fuzzy set theory

3.1.1 Crisp sets versus fuzzy sets

Expert knowledge-based models, or simply expert models or knowledge-based models, are meaningful to the broad range of persons involved in the modelling process because these models allow a clear presentation of the model structure and the inference process (Borsuk et al., 2004). Hence, model interpretability is one of the main features of knowledge-based systems, which can be an excellent basis for collaboration between river managers, scientists and other stakeholders (Casillas et al., 2003). Moreover, neglect of ecological knowledge is still a limiting factor in the application of statistical modelling in ecology and conservation planning (Austin, 2002; Borsuk et al., 2004). A common drawback of most standard modelling approaches is that they cannot incorporate extra information, such as the knowledge of local experts, which is often qualitative and imprecise (Babuška and Verbruggen, 1997).

The fuzzy rule-based models applied in this dissertation allow incorporation of such qualitative knowledge into the modelling process (Casillas et al., 2003). These models are based on fuzzy set theory, which was developed by Zadeh (1965). At the heart of this theory lies the assumption that the world consists of complex systems, which are characterised by fuzzy transitions between different states of the system such as the transition between a nice sunny day and a thunderstorm. As an extension of Boolean logic, fuzzy logic states that a system is not constantly in either state A or state B, but can also be in an intermediary state, described as ‘partly A and partly B’ (Zadeh, 1965).

In a fuzzy rule-based model, each state is described by a fuzzy set, which shows overlapping boundaries with its neighbouring sets. River depth, for example, could be expressed in three fuzzy sets, to which the linguistic values ‘low’, ‘medium’ and ‘high’ could be assigned. When using classical sets with crisp boundaries (hereafter called crisp sets), for instance depths below 1.5 m would be considered ‘low’, depths between 1.5 and 3 m ‘moderate’ and depths higher than 3 m ‘high’. A given depth would either belong to a set (it has a membership degree of one to this set) or it would not (it has a membership degree of 0 to this set). A fuzzy set is described by its membership function, indicating the membership degree for each variable value to this set. As the boundaries between these sets are overlapping, an element can partially belong to a fuzzy set and thus have a membership degree to this set ranging from zero to one. Hence, the linguistic statement ‘the depth is quite low but tending to be moderate’ could be translated into a depth which could have a membership degree of 0.4 to the ‘low’ fuzzy set and of 0.6 to the ‘moderate’ set.

Such assignment corresponds much better with reality than when crisp sets would be used. Specifically, fish would not differentiate between a depth of 1.49 m and a depth of 1.51 m, whereas crisp sets would categorise these depths respectively as ‘low’ and ‘moderate’. Although most habitat suitability models apply the crisp boundary approach, transitions in ecology are not crisp but gradual, resulting in ecological gradients (Cadenasso et al., 2003a; Cadenasso et al., 2003b). Fuzzy logic has proven to be an appropriate modelling technique to deal with these ecological gradients because the boundaries between the classes of the input variables are overlapping in a fuzzy model and thus reflect these gradual transitions between classes (Adriaenssens, 2004; Mouton et al., 2007). Consequently, the fuzzy rule-based models which were applied in this dissertation allow incorporation of the ecological boundary theory (Strayer et al., 2003) in the modelling process.

From a mathematical point of view, a fuzzy set is a function from the domain **X** to the unit interval [0,1] that maps an element x to $A(x)$:

$$A: \mathbf{X} \rightarrow [0,1] : x \rightarrow A(x) \tag{3.1}$$

The value of the membership function A in x is called the membership degree of x to A and ranges from 0 (x does not belong to the fuzzy set) to 1 (x completely belongs to the fuzzy set). If the membership degree is between 0 and 1, x partially belongs to the fuzzy set. A crisp set is a special case of a fuzzy set with membership degrees restricted to {0,1}. In Fig. 3.1, the linguistic values ‘low’, ‘moderate’ and ‘high’ of the linguistic variable depth are defined by membership functions in the domain **D** = [0,5].

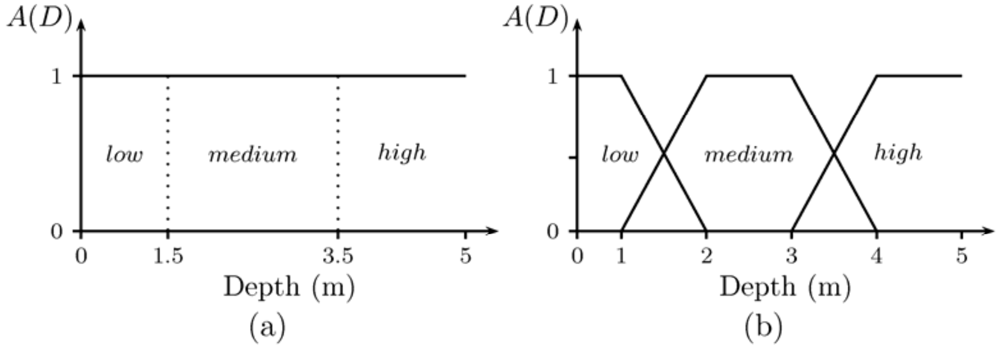


Fig. 3.1. Defenition of the three linguistic values assigned to depth by means of crisp (a) and fuzzy (b) sets.

3.1.2 Mathemathical description of fuzzy sets

Although any function of the form A: $\mathbf{X} \rightarrow [0,1]$ can describe a membership function associated with a fuzzy set A, in this dissertation all membership functions will have trapezial or triangular shapes. A trapezial fuzzy set can be described by four parameters (a_1, a_2, a_3, a_4) and can be defined as:

$$A(x)=\begin{cases}0 & \textit{if } x<a_1\ ,\\ \dfrac{x-a_1}{a_2-a_1} & \textit{if } x\in[a_1,a_2],\\ 1 & \textit{if } x\in[a_2,a_3],\\ \dfrac{a_4-x}{a_4-a_3} & \textit{if } x\in[a_3,a_4],\\ 0 & \textit{if } x>a_4\ .\end{cases} \tag{3.2}$$

The membership degree linearly increases between a_1 and a_2 from 0 to 1, is equal to 1 between a_2 and a_3 and linearly decreases from 1 to 0 between a_3 and a_4. A triangular membership function is obtained when a_2 equals a_3. All membership functions in this dissertation form a fuzzy partition (Ruspini, 1969), which guarantees an interpretable description of the linguistic values (Jin, 2003; Bodenhofer and Bauer, 2005). A family $(A_i)_{i=1}^n$ of n membership functions forms a fuzzy partition of a domain **X** if for each element x of **X** the sum of its n membership degrees to all membership functions equals one:

$$(\forall x\in\mathbf{X})\left(\sum_{i=1}^{n}A_i(x)=1\right)\ . \tag{3.3}$$

The membership functions defining the three linguistic values 'low', 'moderate' and 'high' of the linguistic variable depth in Fig. 3.1 form a fuzzy partition.

Similar to the operations used in classical set theory, three basic operations of fuzzy set theory – intersection, union and complement – can be defined. For two fuzzy sets *A* and *B* defined in a domain X, the intersection of *A* and *B* defined by means of a triangular norm or t-norm *T*:

$$A\cap B(x)=T(A(x),B(x))\ . \tag{3.4}$$

The union of *A* and *B* is defined by means of a triangular co-norm or t-conorm *S* and defined by:

$$A\cup B(x)=S(A(x),B(x))\ . \tag{3.5}$$

Finally, the complement of *A* is defined by:

$$coA(x)=1-A(x)\ . \tag{3.6}$$

Although several definitions have been suggested for t-norms and t-conorms, three definitions are most commonly applied: the minimum t-norm $T_\mathbf{M}$, the product t-norm $T_\mathbf{P}$ and the Lukasiewicz t-norm $T_\mathbf{L}$

$$T_\mathbf{M}(a,b)=\min(a,b) \tag{3.7}$$

Likewise, the product t-norm and the Lukasiewicz t-norm can be defined as:

$$T_\mathbf{P}(a,b)=a\ .\ b\ , \tag{3.8}$$

$$T_\mathbf{L}(a,b)=\max(0,a+b-1)\ . \tag{3.9}$$

and the corresponding t-conorms, the maximum $S_\mathbf{M}$, the algebraic sum $S_\mathbf{P}$ and the Lukasiewicz t-conorm $S_\mathbf{L}$ as:

$$S_\mathbf{M}(a,b)=\max(a,b)\ , \tag{3.10}$$

$$S_\mathbf{P}(a,b)=a+b-a\ .\ b\ , \tag{3.11}$$

$$S_\mathbf{L}(a,b)=\min(1,a+b)\ . \tag{3.12}$$

3.2 Fuzzy rule-based models

3.2.1 Fuzzy rules

The key component of fuzzy rule-based models is the fuzzy rule base, which consists of fuzzy rules of the form

IF antecedent part THEN consequent part.

The fuzzy rule base contains the expert knowledge which links the input variables with the output variable of the model. The antecedent part of the rules defines when the rule holds and the consequent describes the corresponding conclusion of the fuzzy model. A fuzzy habitat suitability model for a fish species could for instance contain the following rule: IF depth is low AND flow velocity is medium AND substrate is coarse THEN the habitat is highly suitable for this species. The knowledge which is incorporated in the model can be described linguistically in the fuzzy rule base (Casillas et al., 2003). The term ‘fuzzy’ is somehow misleading because several authors pointed out that fuzzy systems can approximate any continuous real function on a compact domain to any degree of accuracy (Buckley, 1993; Ying et al., 1999; Perfilieva and Kreinovich, 2002; Campello and do Amaral, 2006).

Two main types of fuzzy rule-based models can be distinguished based on the structure of the rules. In linguistic fuzzy models, both the antecedent and the consequent contain linguistic values, whereas in Takagi-Sugeno models, the antecedent contains linguistic values and the consequent contains a crisp function of the input variables. This dissertation will focus on linguistic fuzzy models. The r rules of a linguistic fuzzy model with m input variables X_l ($l \in L = \{1, \dots, k\}$) and one output variable Y are of the form

$$R_s : \text{IF } X_1 \text{ IS } B^1_{j_{1,s}} \text{ AND} \dots \text{AND } X_m \text{ IS } B^m_{j_{m,s}} \text{ THEN } Y \text{ IS } A_{i_s}$$

where $B^l_{j_{l,s}}$ (resp. A_{i_s}) are linguistic values of variable X_l (resp. Y) in the domain $\mathbf{X}_l$ (resp. $\mathbf{Y}$) ($s \in S = \{1, \dots, r\}$). The input vector is denoted by $\mathbf{x} = (x_1, \dots, x_m)$.

3.2.2 Mamdani-Assilian inference

Mamdani-Assilian models are linguistic fuzzy models which apply t-norm based inference (Assilian, 1974; Mamdani, 1974). For each instance of the input vector x, its membership degrees $B^l_{j_{l,s}}(x_l)$ to the linguistic values in the antecedents of the rules, *i.e.* to the membership functions of each input variable, is determined. In Fig. 3.2 the membership degrees of the input variables x_1 and x_2 ($\mathbf{x} = (x_1, x_2)$) are:

$$B^1_1(x_1) = 0 \qquad B^1_2(x_1) = 0.75 \qquad B^1_3(x_1) = 0.25$$

$$B^2_1(x_2) = 0.33 \qquad B^2_2(x_2) = 0.67$$

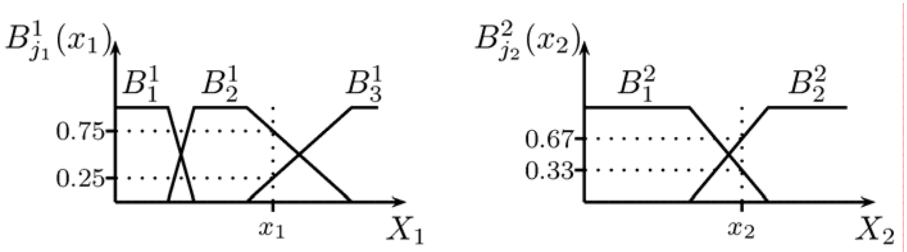


Fig. 3.2. Determining the membership degrees of the model input vector $\mathbf{x} = (x_1, x_2)$ to the linguistic values of the input variables X_1 and X_2.

Next, the fulfilment degrees β_s of the r rules (s $\in$ S = {1, … , r}) are calculated from the membership degrees $B^l_{j_{l,s}}$ (x_l) to the membership functions of each input variable. For the t-norms $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$ this results in:

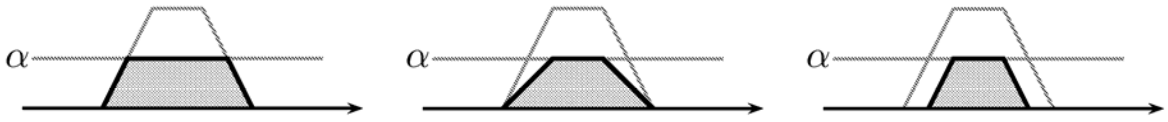
$$\beta_s = \begin{cases} \min\limits_{l=1}^{m} B^l_{j_{l,s}}(x_l) & \text{if } T = T_{\mathbf{M}}\,, \\ \prod\limits_{l=1}^{m} B^l_{j_{l,s}}(x_l) & \text{if } T = T_{\mathbf{P}}\,, \\ \max(\sum\limits_{l=1}^{m} B^l_{j_{l,s}}(x_l) - (m-1), 0) & \text{if } T = T_{\mathbf{L}}\,. \end{cases} \tag{3.13}$$

In the following step, the same t-norm T as the one used for the fulfilment degrees β_s is applied to compute the adapted membership functions $B^{'}_s(y)$ (see Fig. 3.3)

$$A^{'}_{i_s}(y) = \begin{cases} \min(\beta_s, A_{i_s}(y)) & \text{if } T = T_{\mathbf{M}}\,, \\ \beta_s . A_{i_s}(y) & \text{if } T = T_{\mathbf{P}}\,, \\ \max(\beta_s + A_{i_s}(y) - 1, 0) & \text{if } T = T_{\mathbf{L}}\,. \end{cases} \tag{3.14}$$

and the global fuzzy output $A(y)$ is determined as follows:

$$A(y) = \max_{s=1}^{r} A^{'}_{i_s}(y) \tag{3.15}$$



(a) $T_{\mathbf{M}}$ (b) $T_{\mathbf{P}}$ (c) $T_{\mathbf{L}}$

Fig. 3.3. Adapted membership functions (in black) obtained by applying Eq. (3.14) with $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$ to the membership function in grey.

Finally, the fuzzy output is defuzzified, which results in the crisp model output y^*. Two defuzzification methods have been applied most commonly: the Center of Gravity (COG) defuzzification resulting in the crisp model output y^*_{COG} and the Mean of Maxima (MOM) defuzzification resulting in the crisp model output y^*_{MOM} (Kruse et al., 1994)

$$y^*_{COG} = \frac{\int_Y yA(y)dy}{\int_Y A(y)dy} \tag{3.16}$$

$$y^*_{MOM} = \frac{\int_{core(A)} ydy}{\int_{core(A)} dy} \tag{3.17}$$

The core of A contains the set of elements which have the largest degree of membership to a fuzzy set A:

$$core\,(A) = \{x_1 \in \mathbf{X} \mid (\forall\ x_2 \in \mathbf{X})(A(x_2) \leq A(x_1))\}\ . \tag{3.18}$$

If the core of the fuzzy model output A is a set of discrete values, the crisp output y^*_{MOM} is defined as the average of these discrete values.

In practice, Eqs. (3.14)-(3.15) are implemented in a slightly different way. A fulfilment degree α_i is calculated for each linguistic output value A_i based on the fulfilment degrees β_s of the r rules, a fulfilment degree α_i is computed for each linguistic output value A_i,

$$\alpha_i = \max\{\ \beta_s \mid i_s = i\}\ . \tag{3.19}$$

An adapted membership function $A_i^{'}$ is determined for each linguistic output value with the corresponding fulfilment degree α_i

$$A^{'}_{i_s}(y) = \begin{cases} \min(\alpha_i, A_i(y)) & \text{if } T = T_{\mathbf{M}}\ , \\ \alpha_i . A_i(y) & \text{if } T = \mathrm{T}_{\mathbf{P}}\ , \\ & \\ \max(\alpha_i + A_i(y) - 1{,}0) & \text{if } T = \mathrm{T}_{\mathbf{L}}\ . \end{cases} \tag{3.20}$$

and the global fuzzy output $A(y)$ is determined as follows

$$A(y) = \max_{i=1}^{n} A_i^{'}(y) \tag{3.21}$$

with n the number of linguistic values of the output variable. Van Broekhoven et al. (2006) improved the computational efficiency of the COG method by developing a slope based method for the calculation of the crisp model output y^*_{COG}. In this method, the surface described by the fuzzy output is partitioned such that the slope of the fuzzy output is constant within each part and different in two adjacent parts (Van Broekhoven et al., 2006).

In Fig. 3.4, the Mamdani-Assilian inference procedure is illustrated for a model with two input variables, X_1 and X_2 and one output variable Y. The membership functions form a fuzzy partition and describe the linguistic values of all three variables. These values are 'low' and 'high' for X_1 and 'low', 'medium' and 'high' for X_2 and Y. The fulfilment degree β_s of each of the six presented rules is the minimum of the membership degree of x_1 and x_2 to the corresponding linguistic value in the antecedent of the rule. According to this fulfilment degree β_s, the membership functions of the output variable Y in the consequent part of the rules are truncated ($T = T_{\mathbf{M}}$). The union, based on the maximum, of all these truncated fuzzy sets is the global fuzzy output. This output can be defuzzified by the COG defuzzification method, resulting in the crisp model output y^*_{COG}.

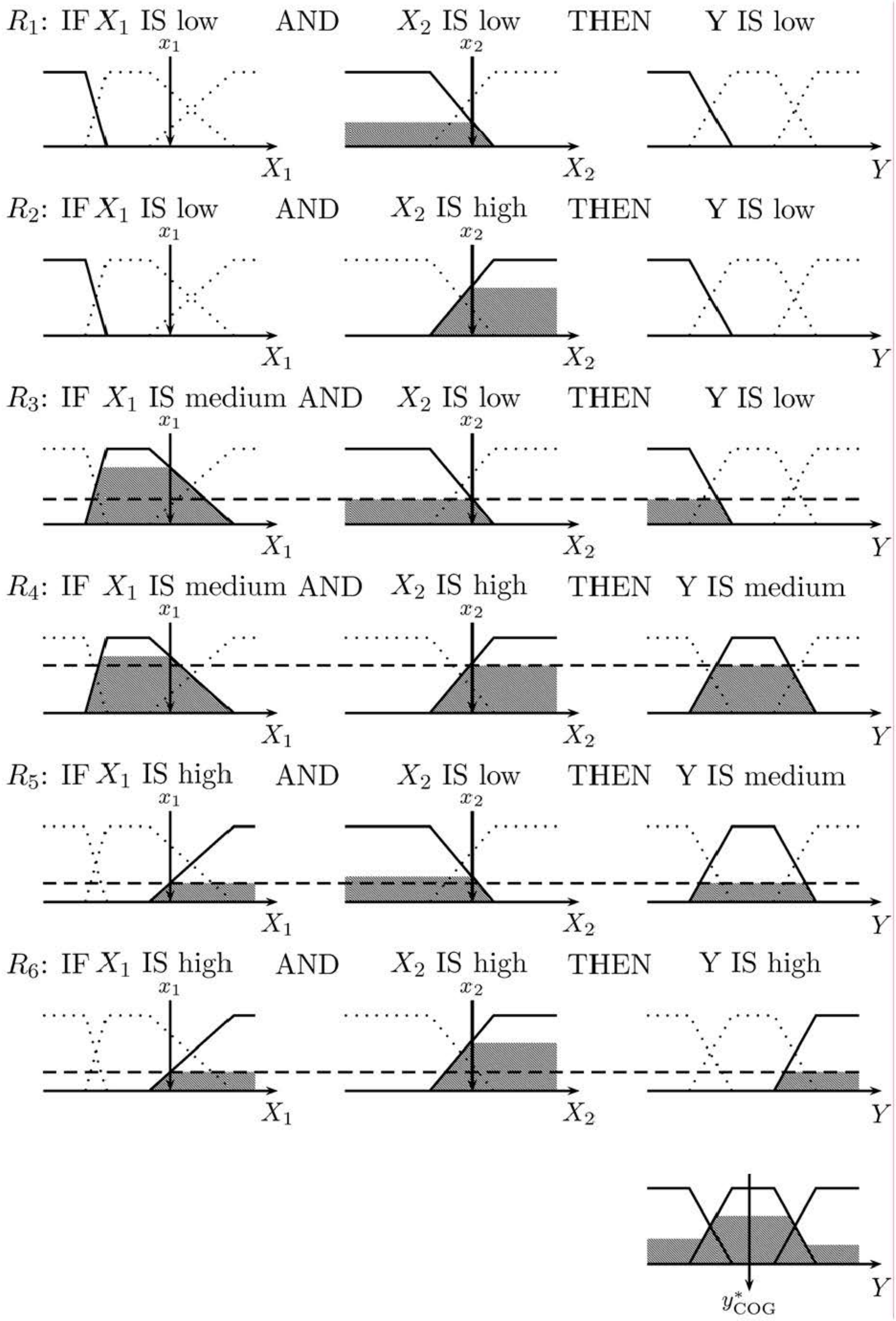


Fig. 3.4. Illustration of Mamdani-Assilian inference ($T = T_M$, COG defuzzification) applied to a model with six rules

3.2.3 Fuzzy ordered classification

In this dissertation, the output of the fuzzy model is calculated by first determining the membership degrees of the crisp input values to the linguistic values of the input variables. Next, the degree of fulfilment is calculated for each rule as the minimum of the fulfilment degrees in its antecedent. Finally, to each linguistic output value a fulfilment degree is assigned equal to the maximum fulfilment degree obtained for all rules containing the linguistic output value under consideration in their consequent. Up to this point, the procedure is the same as the one applied in Mamdani-Assilian models (Assilian, 1974; Mamdani, 1974). However, it is not the purpose of a habitat suitability model to predict a precise numerical value for the occurrence of a given species. Specifically, ecologists, river managers and stakeholders are interested in the magnitude of the abundance and may thus not even trust a model stating an occurrence of, e.g. 77 organisms. Moreover, in this dissertation the output variable was already defined by crisp sets: four abundance sets in Chapter 5 and two crisp sets (present and absent) in Chapters 7 to 9 (Table 3.1). Therefore, a different type of model was applied, more specifically a fuzzy classifier, which implies that the model output of the developed models was fuzzy. Specifically, if n is the number of fuzzy sets of the output variable, the model output y_{model} is a set of n values between zero and one and summing up to one. These values express the degree to which the considered river site is regarded suitable as a habitat for a species. The output is obtained by normalising the fulfilment degrees of the output classes. Note that the output values included in the validation dataset are crisp values (integers). When comparing the fuzzy model outputs with the information in the validation dataset, the membership degrees of the crisp abundance values to the linguistic output values are used.

Table 3.1. The crisp sets defining output variable in the different chapters in this dissertation.

Chapter	Label of output variable	Labels of crisp sets
5	Habitat suitability for the mayfly *Baetis rhodani*, expressed as species abundance	Low, Medium, High, Very high
7	Habitat suitability for European grayling	Absent/Present
8	Habitat suitability for European grayling	Absent/Present
9	Habitat suitability for the caddisfly *Aoteapshyche* spp.	Absent/Present
9	Habitat suitability for Large brown trout	Absent/Present
9	Habitat suitability for Large rainbow trout	Absent/Present

To assess the performance of the developed models, different performance criteria were applied as described in the next chapters. To take into account the fuzzy characteristics of the model output, two performance criteria were applied that can deal with the fuzzy model output: the average deviation (AD) and the adjusted average deviation (aAD), as described in chapter 4. However, most performance criteria are based on the confusion matrix (Chapter 4) and can not deal with the fuzzy model outputs. Therefore, to apply these criteria the model output was assigned to the fuzzy set with the highest fulfilment degree, which allowed comparison of the modelled output with the observed output and calculation of performance measures. If the output variable consists of two sets, this approach is very similar to the defuzzification procedures used in Mamdani-Assilian models.

CHAPTER 4

Fuzzy rule base training

4.1 Introduction

Although linguistic fuzzy rules originally were derived from expert knowledge, more recently various techniques were developed to identify the rules and fuzzy sets from data, such as fuzzy clustering, neural learning methods or genetic algorithms (Guillaume, 2001; Hüllermeier, 2005). These data-driven identification methods for fuzzy models have been developed mostly for Takagi-Sugeno models, because these methods focused on the model's accuracy. Recently, model interpretability also gained importance and several methods have been proposed to obtain a balance between both model accuracy and interpretability. Espinosa and Vandewalle (2000) improved the interpretability of a model obtained by purely data-driven identification, while other authors enhanced the interpretability of accurate fuzzy models (Casillas et al., 2003a) and improved the accuracy of linguistic fuzzy models with a good interpretability (Casillas et al., 2003b). In this dissertation, not only the accuracy of interpretable linguistic models will be improved, but a method will be presented which guarantees that the increase in accuracy is also ecologically relevant.

4.2 Fuzzy sets optimisation

The parameters of the membership functions corresponding to the fuzzy sets of the input variables have often been derived from expert knowledge. However, if a fuzzy set of an input variable contains very few training instances, rules which involve this fuzzy set will be trained inadequately. Therefore, in this dissertation a uniform distribution of the input variables over the fuzzy sets was suggested to generate reliable rule bases. The Shannon–Weaver entropy (Shannon and Weaver, 1963) quantified this uniformity and was applied to optimise the parameters of the membership functions of the input variables. The fuzzy sets were converted into crisp ones whose boundaries were the points having a membership degree of 0.5 to the corresponding fuzzy set.

The entropy E is given by (convention $0 \log_2 0 = 0$):

$$E = -\frac{1}{\log_2 n}\sum_{i=1}^{n} p_i \log_2 p_i \tag{4.1}$$

where n is the number of classes and p_i is the proportion of data points belonging to class i. The algorithm for parameter optimisation starts with n equal to 2, and then the parameters of the fuzzy sets are adjusted in steps of $r/2s$ with r the range of the variable of which the fuzzy sets are optimised, and s the fixed stepsize. For each variable, parameter optimisation started by creating two crisp sets with boundary at r/n with n the number of sets, which equals two. The crisp sets were transformed into fuzzy sets by setting the parameters of each set m, $a_{m,t}$, $b_{m,t}$, $c_{m,t}$ and $d_{m,t}$ (Fig. 1) as follows:

$$c_{m,t} = a_{m+1,t} = (m+1)\cdot\frac{r}{n} - t\cdot\frac{r}{2s} \quad , \tag{4.2}$$

$$d_{m,t} = b_{m+1,t} = (m+1)\cdot\frac{r}{n} + t\cdot\frac{r}{2s} \quad , \tag{4.3}$$

with t indicating the iteration of the sets optimisation and thus $t = 1$ in this situation. Next, the entropy $E_{n,t}$ of this fuzzy set configuration was calculated as described in Eq. (4.1) and this entropy was set to E_{best}. This entropy E_{best} is compared with the entropy fixed threshold E_{thres} and the algorithm was terminated if $E_{best} > E_{thres}$.

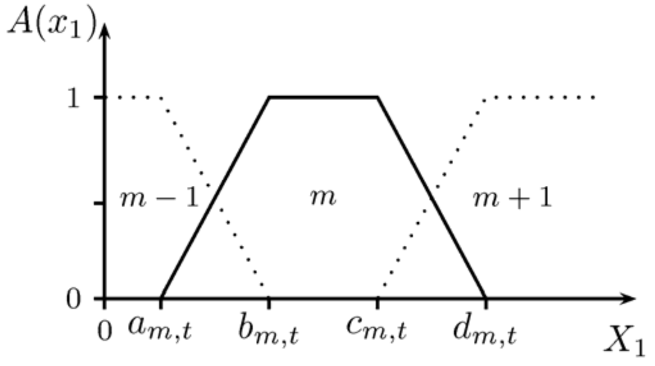


Fig. 1. The parameters of the fuzzy set m of variable X_1.

If $E_{best} < E_{thres}$, the algorithm searched for the fuzzy set of which an expansion of the set boundaries could lead to the greatest increase in entropy. For each set m, the upper boundary was expanded as follows:

$$c_{m,t+1} = a_{m+1,t+1} = (m+1)\cdot\frac{r}{n} - (t+1)\cdot\frac{r}{2s} \quad , \tag{4.4}$$

$$d_{m,t+1} = b_{m+1,t+1} = (m+1)\cdot\frac{r}{n} + (t+1)\cdot\frac{r}{2s} \quad . \tag{4.5}$$

The entropy of this new fuzzy set configuration in iteration t+1, $E_{n,t+1}$, was calculated and compared to E_{best}. If $E_{n,t+1} > E_{best}$, the algorithm continued with this new configuration and the upper boundary of the set m was further adjusted according to Eqs. (4.4) and (4.5). The last adjustment of the fuzzy set was cancelled and the algorithm continued with adjusting the upper boundaries of the next fuzzy set m+1 if $E_{n,t+1} < E_{best}$,.

To avoid that the algorithm would generate erroneous fuzzy sets in which the total membership degree of a variable value would exceed one, the boundary adjustment of the fuzzy sets in each iteration t was limited as follows:

$$(t+1)\cdot\frac{r}{2s}\leq\min(\min(c_{m,1}-b_{m,1},c_{m+1,1}-b_{m+1,1}),\min(c_{m,t}-b_{m,t},c_{m+1,t}-b_{m+1,t}))\quad . \tag{4.6}$$

If the boundary adjustment of fuzzy set m reached this limit and the entropy was still lower than E_{thres}, the set was split into two symmetric sets. This split was created in the same way as the first split of the total variable range into two fuzzy sets, but the variable range was now replaced by the range of the fuzzy set m. The entropy of this new set configuration, $E_{n+1,1}$, was calculated and compared to E_{best}. If $E_{n+1,1} < E_{best}$, the algorithm did not split the sets and continued with the next fuzzy set. If $E_{n+1,1} > E_{best}$, the algorithm restarted from this new configuration. If an entropy which exceeded E_{thres} was obtained during sets optimisation, the boundaries of the remaining sets were still optimised, but without splitting the sets.

The applied fuzzy sets optimisation method ensured that the distribution of the training data instances over the fuzzy sets was optimal. In a situation with n fuzzy sets, sets which contained less than $100/n$ % of the data would be expanded as far as possible, whereas sets which contain more than $100/n$ % of the data would be reduced to or split into smaller sets. This method avoided that empty or poorly represented sets were included in the model and increased model efficiency by deleting redundant sets. Several authors have suggested similar approaches and demonstrated that such methods significantly improve model performance (Casillas et al., 2003a; Casillas et al., 2003b). However, a more uniform distribution of the input data over the fuzzy sets does not guarantee that each fuzzy rule is represented uniformly in the input data. Water depth values, for instance, can be uniformly distributed over a ‘low’ and a ‘high’ fuzzy set, but this does not imply that many sampling sites with a high flow velocity and a high depth are represented in the training set.

4.3 Hill-climbing algorithms for rule base training

The term ‘hill-climbing’ implies an iterative improvement technique, and thus describes both maximisation and minimisation problems. The technique is applied to a single solution, the current solution, in the search space. During each iteration, a new solution is selected from the neighbourhood of the current solution. If that solution provides a better value in light of the evaluation function, the new solution becomes the current solution. Otherwise, some other neighbour is selected and tested against the current point. The method terminates if no further improvement is possible. Hill-climbing methods are often started from a large variety of different starting solutions because these methods can only provide locally optimal values, and these values depend on the selection of the starting point. In this dissertation, the initial points are chosen at random, and the algorithm is stopped if the best solution is found for the fifth time. For problems with many local optima, particularly those where these optima have large basins of attraction, it’s often very difficult to locate a globally optimal solution. However, several authors have shown that there is no way to choose a single search method that can serve well in every case (Fogel and Ghozeil, 1997; Wolpert and Macready, 1997; Michalewicz and Fogel, 2000), which is referred to as the 'no free lunch theorem'.

There are a few versions of hill-climbing algorithms, which differ mainly in the way a new solution is selected for comparison with the current solution. In this dissertation, the steepest ascent hill-climbing algorithm is applied. Initially, all possible neighbours v_n of the current solution are considered, and the one v_n that returns the best model performance is selected to compete with the current solution, v_c. If the model performance of v_c is worse than the performance of v_n, then the new solution v_n becomes the current solution. Otherwise, no local improvement is possible and the algorithm has reached a local or a global optimum. In such a case, the next iteration of the algorithm is executed with a new current solution selected at random (Michalewicz and Fogel, 2000).

To generate a reliable habitat suitability model, the consequents of the fuzzy rules in this dissertation were optimised using a nearest ascent hill-climbing algorithm. First, the fuzzy sets were optimised as described in Section 4.2. Once the sets were optimised, they did not change during the further training procedure. Starting from these fixed fuzzy sets and a rule base with randomly selected rule consequents, the linguistic value in the consequent of one randomly selected rule was changed into its neighbouring linguistic value and the impact on model performance was calculated. If model performance increased, the algorithm continued with the adjusted rule base, if not, it continued with the original one. If a linguistic term had two neighbouring linguistic terms (e.g. the linguistic term ‘moderate’ may have the two neighbouring linguistic terms ‘low’ and ‘high’), the impact of both neighbouring linguistic terms on model performance was calculated and compared. To indicate the robustness of the optimisation results, *n*-fold cross-validation was applied. First, *n* partitions were constructed by randomising the original dataset and assigning each data point to one partition without replacement, such that each partition contained 100/*n* % of the total number of data points of the dataset. Ten sets of both a training and a test fold were then created by identifying one partition as the test fold and by grouping the *n*-1 other partitions into the training fold. This procedure resulted in *n* different training and test folds. The species prevalence (i.e. the frequency of occurrence) and the fuzzy set configuration was constant for all partitions and thus for all training and test folds.

Different models were trained based on different performance measures as described in the next sections. Different training scenarios were created to compare the results of optimisation based on the different performance measures (Table 4.2). Each training iteration was stopped when no further increase of the performance measure on the test fold was observed. Each training iteration was repeated and the obtained rule base was compared to each rule base obtained in previous iteration steps. The resulting rule base similarity indicated the percentage of rule consequents that was identical for two rule bases. If the rule base with the highest performance on the test fold was obtained 5 times, this rule base was selected as the final rule base and training continued on another fold as in the following algorithm:

Algorithm 4.1: Training algorithm

```
$t \leftarrow 0$
for each fold do
    $stop \leftarrow 0$
    while $stop < 5$ do
        $t \leftarrow t + 1$
        Train rule base based on training fold
        $Perftest_t \leftarrow$ the performance of the resulting rule base, $RB_t$, on the test fold
        if $t = 1$ then
            Maximal $Perftest \leftarrow Perftest_t$
            Maximal $RB \leftarrow RB_t$
        else if $Perftest_t >$ Maximal $Perftest$ then
            Maximal $Perftest \leftarrow Perftest_t$
            Maximal $RB \leftarrow RB_t$
            $stop \leftarrow 0$
        else if $Perftest_t =$ Maximal $Perftest$ then
            if similarity of $RB_t$ and Maximal $RB$ = 100 % then
                $stop \leftarrow stop + 1$
            end
        end
    end
end
```

In this work, the fuzzy rule base contained fuzzy rules representing each possible combination of input variable sets. However, not every combination of input variable sets was present in the studied stretch. To obtain an indication of which rules were relevant, the fuzzy sets were turned into crisp ones by assuming that an input value belongs not to a set if its membership degree to this set is < 0.5. Each data point could be assigned to one environmental condition, resulting in the distribution of the data points over the 'crisp' environmental conditions described by the rules. The distribution also gives an indication of the usefulness of the obtained rule base over a range of environmental conditions that can be found in the study area. The similarity between the rule base of model A and model B was calculated as the % *CCI*, assuming that the rule base of model A equals the observed values and the one of model B the predicted output.

4.4 Performance criteria

The key component of the model training and validation procedures is the performance criterion which evaluates the model performance. Performance criteria can deal with either continuous or discrete model outputs, or with both. If a model generates discrete predictions, these outputs can be summarized in a confusion matrix (Fielding and Bell, 1997; Manel et al., 2001) which compares the model predictions to the observations (Table 4.1). No individual element of the matrix provides a synoptic view of overall model performance, while conventional statistics of association on these confusion matrices are inappropriate for assessing model performance (Manel et al., 2001). Therefore, several performance criteria have been derived from this confusion matrix, including overall predictive accuracy or the percentage of correctly classified instances (*CCI*; Buckland and Elston, 1993; Fielding and Bell, 1997), sensitivity, specificity, the normalized mutual information statistic (*NMI*; Forbes, 1995), *Kappa* (Cohen, 1960), the odds ratio (Fielding and Bell, 1997) and the true skill statistic (*TSS*; Allouche et al., 2006). The latter two criteria range from -1 to 1, whereas all other criteria range from 0, where models are completely inaccurate, to 1, where presence-absence is perfectly predicted.

Table 4.1. The confusion matrix as a basis for evaluation of observed data. The table cross-tabulates observed values against predicted values: true-positives, *a*; false-positives, *b*; false-negatives, *c*; true-negative values, *d*.

		Observed	
		Present	Absent
Predicted	Present	*a*	*b*
	Absent	*c*	*d*

Table 4.2. Measures of predictive accuracy calculated from a 2x2 error matrix (Table 4.1). The percentage of Correctly Classified Instances (*CCI*) is the rate of correctly classified cells. Sensitivity (*Sn*) is the probability that the model will correctly classify a presence. Specificity (*Sp*) is the probability that the model will correctly classify an absence. *NMI* quantifies the information included in the model predictions compared to that included in the observations. The *Kappa* statistic and *TSS* normalise the overall accuracy by the accuracy that might have occurred by chance alone. In all formulae $n = a + b + c + d$.

Measure	Formula	
CCI	$\frac{a+d}{n}$	(4.7)
Sn	$\frac{a}{a+c}$	(4.8)
Sp	$\frac{d}{b+d}$	(4.9)
NMI	$1-\frac{-a\ln(a)-b\ln(b)-c\ln(c)-d\ln(d)+(a+b)\ln(a+b)+(c+d)\ln(c+d)}{n\ln(n)+(a+c)\ln(a+c)+(b+d)\ln(b+d)}$	(4.10)
Kappa	$\frac{\left(\frac{a+d}{n}\right)-\frac{(a+b)(a+c)+(c+d)(d+b)}{n^2}}{1-\frac{(a+b)(a+c)+(c+d)(d+b)}{n^2}}$	(4.11)
Odds ratio	$\frac{ad}{cb}$	(4.12)
TSS	$Sn + Sp - 1$	(4.13)

4.5 Selection of input variables

The presented knowledge extraction method can also be applied to select the optimal input variables of the fuzzy model. Such variable selection not only decreases model complexity, but also reduces computational and monitoring efforts (D'heygere et al., 2003; D'heyere et al., 2006; Gabriels et al., 2007). The variable selection method presented in this dissertation, first estimates the contribution of each variable to the global model performance. The contribution of an input variable X_i to a model with q input variables is computed according to Algorithm 4.2.

All variables were then ranked based on their $q!$ rankings. Specifically, the variable with the highest marginal contribution was the variable which was ranked the most often as the variable with the highest contribution.

Finally, the variable contribution method was applied to select the optimal input variables for the habitat suitability model. This selection is based on both model performance, quantified by a performance criterion such as the percentage of correctly classified instances (*CCI*) or *Kappa* (Fielding and Bell, 1997), and on the Akaike information criterion *AIC* (Akaike, 1974). The latter measure not only considers model performance, but also model complexity and thus avoids overfitting of the training data by the model. Starting from a model A which consisted of m variables, this model was first optimised with the hill-climbing algorithm described earlier. Then, input variables were selected as described in Algorithm 4.3.

Algorithm 4.2. Computing the contribution of each variable to the model

```
$c = 0$
initialise $P$, containing the $q!$ possible permutations $P_c$ of $q$ input variables
while $c < q!$
        $c = c + 1$
        create a model $A$ containing all $q$ input variables
        calculate the performance of model $A$, $perf(A)$
        $j = 0$
        $cont_c = 0$
        while the number of variables in model $A$ is greater than one
                $j = j + 1$
                create model $A_j$ by removing the first variable of $P_c$, $X_{i,j}$, from the model $A$
                remove the first variable of $P_c$, $X_{i,j}$, from $P_c$
                calculate the performance of model $A_j$, $perf(A_j)$
                the marginal contribution of $X_{i,j}$ to model $A$, $cont_{i,j,c}$, is given by
                        $cont_{i,j,c} = 1 - cont_c - \frac{perf(A_j)}{perf(A)}$
                $cont_c = cont_c + cont_{i,j,c}$
        end
        for each variable $X_i$ of the $q$ input variables
        assign a ranking $rank_{i,c}$ to $X_i$ based on its marginal contribution to model $A$
end
```

The variable selection method was a combination of a step-forward and a step-backward procedure. Step-forward variable selection starts from a model which contains one variable, and then expands this model by adding the other variables one by one (Gabriels et al., 2007). A step-backward selection method starts from a model which includes all variables, and then reduces this model by removing variables one by one. In both approaches, the contribution of each variable is the difference between the performances of the model with and of the model without this variable. If the input variables are not independent of each other, the step-forward variable contribution to the total model performance differs from the step-backward contribution. Specifically, the step-forward procedure considers different correlations between variables than the correlations which are included in the step-backward approach. Therefore, in this dissertation, the variable selection method applied was a combination of both step-forward and step-backward procedures. Although this approach is computationally more expensive, it considers all possible correlations between variables and thus may generate consistent results.

The starting point of the algorithm may vary based on the available data or expert knowledge. However, the algorithm has been developed such that the starting point does not affect the final result of the algorithm. It should be noted that the highest marginal increase of performance after at random addition of a variable may be a negative value, but this does not affect the result of the algorithm. This variable selection approach was only applied in Chapter 5 for two reasons. First the approach is computationally expensive and thus less appropriate for the larger datasets which are applied in Chapters 7-9. Moreover, in contrast to the dataset applied in Chapter 5, the latter datasets contained only a limited number of input variables, and thus variable selection was less necessary.

Algorithm 4.3. Selection of the input variables

```
for each variable $X_i$ which is not in model $A$
    create a new model $A_i$ by adding $X_i$ to model $A$
    optimise model $A_i$
    compute the performance of model $A_i$
end
select the model $A_i$ with the highest marginal increase of performance*
calculate the contribution of the variables of $A_i$ to $perf(A_i)$ (Algorithm 1)
if $perf(A_i) > perf(A)$ and $AIC(A_i) < AIC(A)$
    if the contribution of the variable $X_i$ last added to the model is the lowest contribution
        restart the algorithm with model $A$ = model $A_i$
    else
        create model $A_i'$ by removing the variable with the lowest contribution
        optimise model $A_i'$
        restart the algorithm with model $A$ = model $A_i'$
    end
else
    if the contribution of the variable $X_i$ last added to the model is the lowest contribution
        stop the algorithm; model $A$ contains the most significant variables
    else
        create model $A_i'$ by removing the variable with the lowest contribution
        optimise model $A_i'$
        if $perf(A_i') > perf(A)$ and $AIC(A_i') < AIC(A)$
            restart the algorithm with model $A$ = model $A_i'$
        else
            stop the algorithm;
            model $A$ contains the most significant variables
        end
    end
end
```

* a decrease of model performance is treated as a negative increase of performance

CHAPTER 5

Knowledge-based versus data-driven models

5.1 Introduction

In this chapter, a fuzzy habitat suitability model for macroinvertebrates in rivers is developed at the mesohabitat scale level. The ecological expert knowledge needed for this model is derived in two different ways in order to compare expert knowledge derived from literature with knowledge derived from data using a hill-climbing algorithm. Habitat models were generated for the mayfly *Baetis rhodani* Pictet in the Zwalm river basin (Belgium), which is an indicator species for a good ecological water quality. The hypothesis tested hereby was that data-driven models outperform expert knowledge-based models. Although a case study is presented of the larvae of the mayfly *Baetis rhodani* Pictet in the Zwalm river, the model could be applied on any other aquatic species and on any other river. The results in this chapter may support river managers to assess bottlenecks in the river basin and thus to efficiently allocate small scale restoration efforts.

5.2 Material and methods

5.2.1 Study area

The study area includes the Zwalm river basin (03° 43’ E; 50° 51’ N) which is a 116.5 km sub-basin of the Scheldt river basin in Flanders, the northern part of Belgium (Fig. 5.1). The mean annual flow at the mouth of this 22 km long river is 1.21 m /s, with the mean summer low flow being 0.53 m /s and the mean high flow 2.43 m /s. Most of the headwaters in the Zwalm river basin are colonised by very rare fish species and several vulnerable macro-invertebrates (Dedecker et al., 2004).

Chemical water quality in the Zwalm river basin improved over the last years, due to investments in sewer systems and wastewater treatment plants (VMM, 2007). However, whereas physical habitat quality is still excellent in the headwaters, it ranges from moderate to very poor in the inhabited parts of the river basin due to flood control weirs, straightened river channels and artificial embankments. Specifically, these flood control weirs obstruct invertebrate and fish migration and are one of the most important ecological problems within the river basin (Dedecker et al., 2004). Many restoration measures have been proposed and planned by the Flemish government (Belconsulting, 2006), but the possible impact of these measures on river biology was never investigated. This emphasises the need for reliable predictive tools providing decision support for integrated river management.

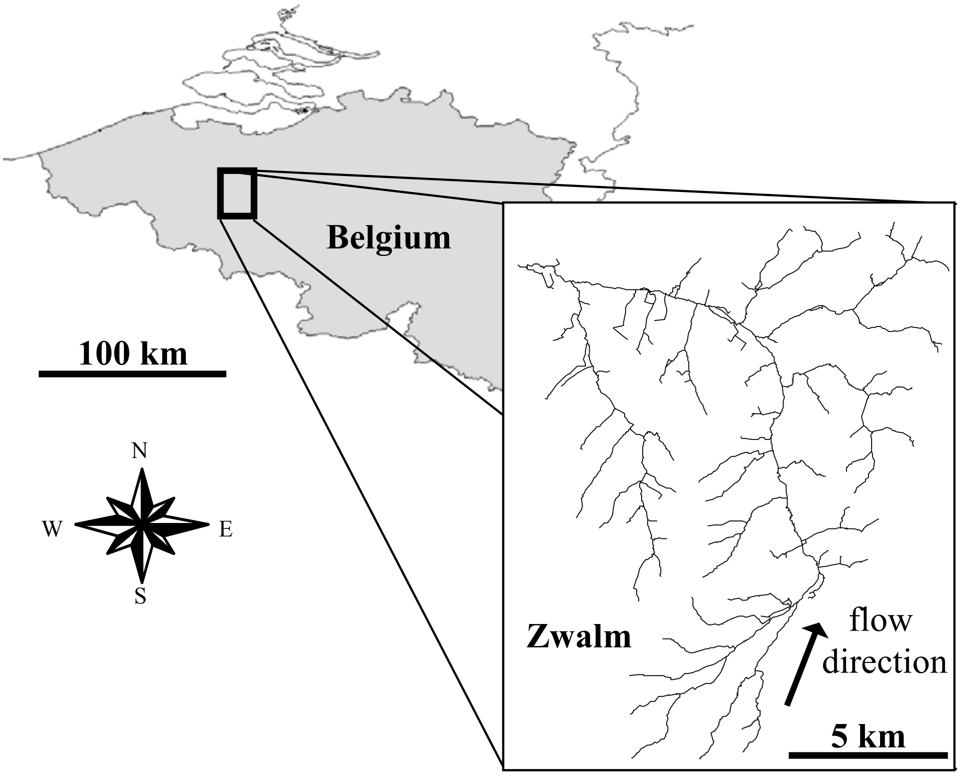


Fig. 5.1. Location of the Zwalm River basin in Flanders, Belgium.

5.2.2 Data Collection

Biological data were collected in the Zwalm river at 323 sites during 5 consecutive years between August and September (2000-2004) to avoid seasonal bias. The macroinvertebrate population was assessed in 10 m of the present habitat at each site by means of 5 minutes kick sampling, using a standard hand net with mesh size 500 μm (IBN, 1984) and by in situ exposure of artificial substrates (De Pauw and Vanhooren, 1983). All the collected material was transferred to buckets with cover and afterwards samples were sieved and organisms sorted in the laboratory. After separation, *Baetis rhodani* (Pictet, 1843) larvae were identified under a stereoscopic dissection microscope (magnification 10-50 times) and abundances were determined. These abundances were log(*abundance* +1) transformed and were assumed to indicate the habitat suitability for *B. rhodani* larvae.

Structural and physical variables were measured to describe the different mesohabitats (Table 5.1). Flow velocity was determined using a propeller flow velocity meter (Höntzsch ZS25GFE). For each 10 m stretch, flow velocity measurements were performed at 40 % of depth on 15 points, divided over 5 transects. Each transect consisted of 3 equidistant points, forming a uniform grid. The dominating substrate was visually assessed and expressed in 4 classes. Water level and width were recorded with a measuring tape while field measurements were performed for dissolved oxygen (OXI 330/SET). Water samples were taken and analysed spectrophotometrically in the laboratory to assess the concentrations of ammonium, nitrate and phosphate.

5.2.3 Fuzzy rule based modelling

The fuzzy rule-based modelling and the variable selection procedure were applied as described in Chapter 4. Adriaenssens (2004) derived an **expert knowledge base from literature**, consisting of fuzzy sets and fuzzy rules. The variables included in this knowledge base were stream width, flow velocity, conductivity (C), and concentrations of ammonium (A), nitrate (N) and phosphate (P). Four different models were created by combining each chemical variable (A, C, N and P) with the two physical variables flow velocity and stream width.

To generate a reliable **knowledge base derived from data**, the consequents of the fuzzy rules were optimised using a nearest ascent hill-climbing algorithm as described in Chapter 4. Three-fold cross-validation was applied to indicate the robustness of the optimisation results. The folds were constructed by randomising the original data set and assigning each data point to one fold without replacement. The species prevalence (i.e. the frequency of occurrence) was constant for all three folds and equal to the prevalence of the original dataset (0.31 = 99/319).

First, the same fuzzy sets were used as those of the literature expert knowledge approach (Table 5.1). However, if a fuzzy set of an input variable contains very few training instances, rules which apply to this fuzzy set will be trained inadequately. Consequently, before rule base optimisation, the fuzzy sets were optimised based on the Shannon-Weaver entropy (Shannon and Weaver, 1963) to create a uniform distribution of the input variables over the fuzzy sets as described in Chapter 4. Since Cohen's *Kappa* (Cohen, 1960) was assumed to be an appropriate performance measure for rule base optimisation, it was applied in this chapter. To further optimise the data-driven model, the impact of different variables on model performance was analysed. Specifically, the importance of each variable was expressed as the average marginal contribution of each variable to the total model output as described in Chapter 4.

Table 5.1. Measured variables at each sampled river stretch, the linguistic values assigned to the input variables of the habitat suitability models and their corresponding fuzzy sets. The expert knowledge-based fuzzy set parameters were derived by Adriaenssens (2004)

Variable	Unit	Linguistic value	Expert knowledge-based fuzzy set parameters	Data-driven fuzzy set parameters
Flow velocity	m.s^{-1}	Low	(0.00,0.00,0.00,0.25)	(0.00,0.00,0.09,0.46)
		Medium	(0.00,0.25,0.25,0.50)	(0.09,0.46,0.55,0.67)
		High	(0.25,0.50,1.30,1.30)	(0.55,0.67,1.22,1.40)
Dominating substrate	4 classes (from 1 = pebble to 4 = loam/clay)	Low	/	(0,1,1,2)
		Medium	/	(1,2,2,3)
		High	/	(2,3,3,4)
		Very high	/	(3,4,4,5)
Depth	m	Low	/	(0.00,0.00,0.22,0.55)
		Medium	/	(0.22,0.55,0.74,0.91)
		High	/	(0.74,0.91,1.65,1.82)
Width	m	Low	(0.0,0.0,0.0,2.0)	(0.00,0.00,0.63,1.72)
		Medium	(0.0,2.0,2.0,4.0)	(0.63,1.72,2.07,3.17)
		High	(2.0,4.0,4.0,6.0)	(2.07,3.17,4.95,6.05)
		Very high	(4.0,6.0,11.0,11.0)	(4.95,6.05,11.00,11.55)
Ammonium	mg NH_4^+-N/l	Very low	(0.00,0.00,0.00,0.10)	/
		Low	(0.00,0.10,0.10,0.15)	(0.00,0.00,0.63,1.63)
		Medium	(0.10,0.15,4.00,5.00)	(0.63,1.63,2.25,2.75)
		High	(4.00,5.00,8.00,10.00)	(2.25,2.75,5.00,5.50)
		Very high	(8.00,10.00,30.00,30.00)	/
Nitrate	mg NO_3^--N/l	Very low	(0.00,0.00,0.00,0.15)	/
		Low	(0.00,0.15,0.15,0.30)	(0.00,0.00,2.80,4.38)
		Medium	(0.15,0.30,0.30,0.40)	(2.80,4.38,6.32,9.48)
		High	(0.30,0.40,0.40,0.45)	(6.32,9.49,15.81,17.39)
		Very high	(0.40,0.45,20,20)	/
Phosphate	mg PO_4^{3-}-P/l	Very low	(0.00,0.00,0.00,0.10)	/
		Low	(0.00,0.10,0.10,0.15)	(0.00,0.00,0.28,0.72)
		Medium	(0.10,0.15,4.00,5.00)	(0.28,0.72,0.99,1.21)
		High	(4.00,5.00,8.00,10.00)	(0.99,1.21,2.2,2.42)
		Very high	(8.00,10.00,30.00,30.00)	/
Conductivity	µS/cm	Very low	(0,0,150,250)	/
		Low	(150,250,450,550)	(0,0,340,1019)
		Medium	(450,550,750,850)	/
		High	(750,850,1050,1150)	(340,1019,1359,1699)
		Very high	(1050,1150,2880,2880)	/
Dissolved oxygen	mg O_2/l	Low	/	(0.0,0.0,5.4,6.6)
		High	/	(5.4,6.6,12.0,12.6)
Habitat suitability	log(abundance +1)	Low	(0.00,0.00,0.00,0.48)	(0.00,0.00,0.25,0.65)
		Medium	(0.00,0.48,0.48,0.78)	(0.25,0.65,0.90,1.10)
		High	(0.48,0.78,1.04,1.32)	(0.90,1.10,2.00,2.20)
		Very high	(1.04,1.32,2.01,2.01)	/

5.3 Results

Fuzzy rule-based models were constructed based on the expert knowledge rules described by Adriaenssens (2004). Data-driven fuzzy models were generated using the same variables. Each model was labelled according to the variables included in the model: depth (D), width (W), flow velocity (V), dominating substrate (S), conductivity (C), oxygen (O), nitrate (N), phosphorus (P) and ammonia (A) concentration. Consequently, model WVC, for instance, included the variables width, flow velocity and conductivity. The performance of both model types was compared by calculating the *CCI* and *Kappa* of all models (Fig. 5.2). All data-driven models performed better than their expert knowledge-based opponents for both *CCI* and *Kappa*. Moreover, the data-driven models contained fewer degrees of freedom than their expert knowledge-based opponents (Table 5.2).

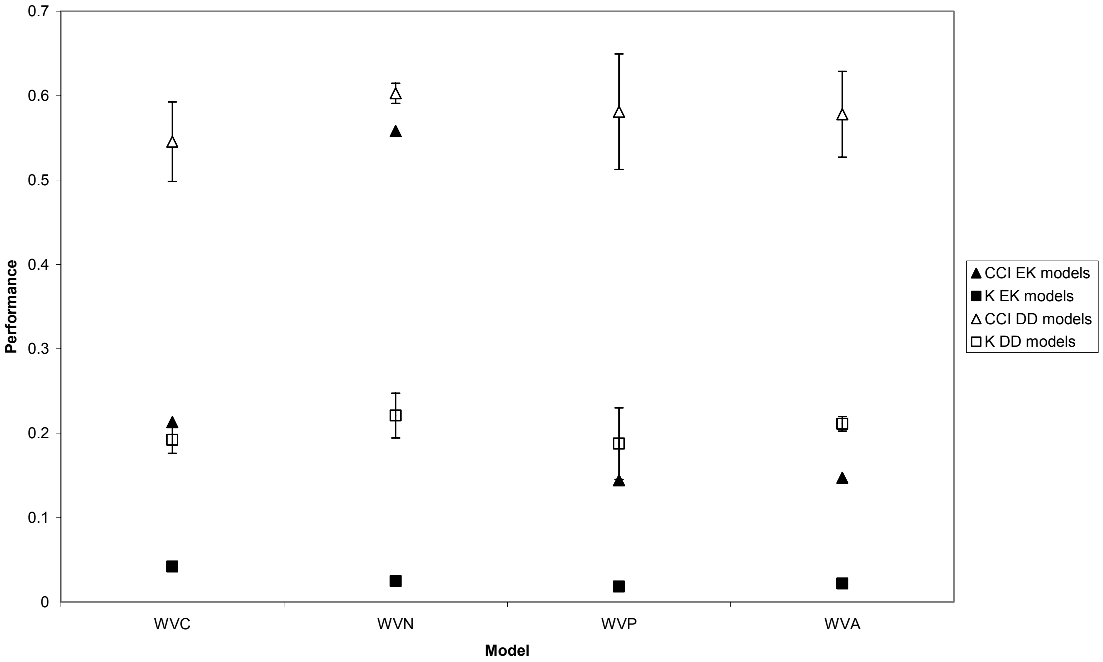


Fig. 5.2. Performance of the four expert knowledge-based (EK = expert knowledge; solid black symbols) and the four data-driven (DD = data-driven; white symbols) models quantified by the percentage of correctly classified instances (*CCI*; triangular symbols) and *Kappa* (*K*; square symbols). Values of these performance criteria were averaged over the three folds for the data-driven models. Each model was labelled according to the variables included in the model: depth (D), width (W), flow velocity (V), dominating substrate (S), conductivity (C), oxygen (O), nitrate (N), phosphorus (P) and ammonia (A) concentration. Consequently, model WVC, for instance, included the variables width, flow velocity and conductivity.

Table 5.2. Proportional contribution of each variable to the total model performance, quantified by *Kappa* and the ranking of each variable, based on its respective contribution. Variable contributions to the data-driven models were averaged over the three folds. EK = Expert knowledge-based model; DD = data-driven model; W = river width; V = flow velocity; A = ammonium concentration; C = conductivity; N = nitrate concentration; P = orthophosphate concentration; WVX = model incorporating the variables W, V and X with X = A, C, N or P.

Model	W	V	A	C	N	P	Degrees of freedom
EK WVA	0.1066 3	0.6959 1	0.1975 2	-	-	-	60
DD WVA	0.232 ± 0.071 2	0.543 ± 0.136 1	0.225 ± 0.065 3	-	-	-	36
EK WVC	0.086 3	0.561 1	-	0.202 2	-	-	60
DD WVC	0.300 ± 0.108 2	0.649 ± 0.135 1	-	0.051 ± 0.028 3	-	-	24
EK WVN	0.026 3	0.369 2	-	-	0.605 1	-	60
DD WVN	0.202 ± 0.107 3	0.549 ± 0.156 1	-	-	0.249 ± 0.149 2	-	36
EK WVP	0.101 3	0.719 1	-	-	-	0.180 2	60
DD WVP	0.173 ± 0.057 3	0.457 ± 0.016 1	-	-	-	0.370 ± 0.063 2	36

The impact of the different variables on the final model performance was quantified by the variable importance and the average ranking of each variable (Table 5.2). The results show that flow velocity is the most important variable in both expert knowledge-based and data-driven models. Although all expert knowledge-based models categorise river width as the least important variable, 75 % of the data-driven models agree that width is the second most important variable for the habitat suitability of mayflies. However, this result is only substantial for the model incorporating conductivity, which indicates that in general, both expert knowledge-based and data-driven models rank the variables similarly.

The purely data-driven model development approach selected flow velocity as the most important variable for the habitat suitability of mayflies (Fig. 5.3). Models were created by combining each of the resulting variables with flow velocity and the performance of these models was analysed (Figs. 5.3-5.4). This indicated that river width also significantly influenced mayfly habitat suitability. Further expansion of the models revealed that first adding the river bed substrate (Fig. 5.5) and then the nitrate concentration of the river substantially increased model performance (Fig. 5.6). However, the model which incorporated the nitrate concentration showed higher AIC values than the model which only contained flow velocity, river width and river bed substrate (Table 5.3). Consequently, no further variables were added to the model. During the whole model development process, the marginal increase in model performance decreased while adding variables to the model.

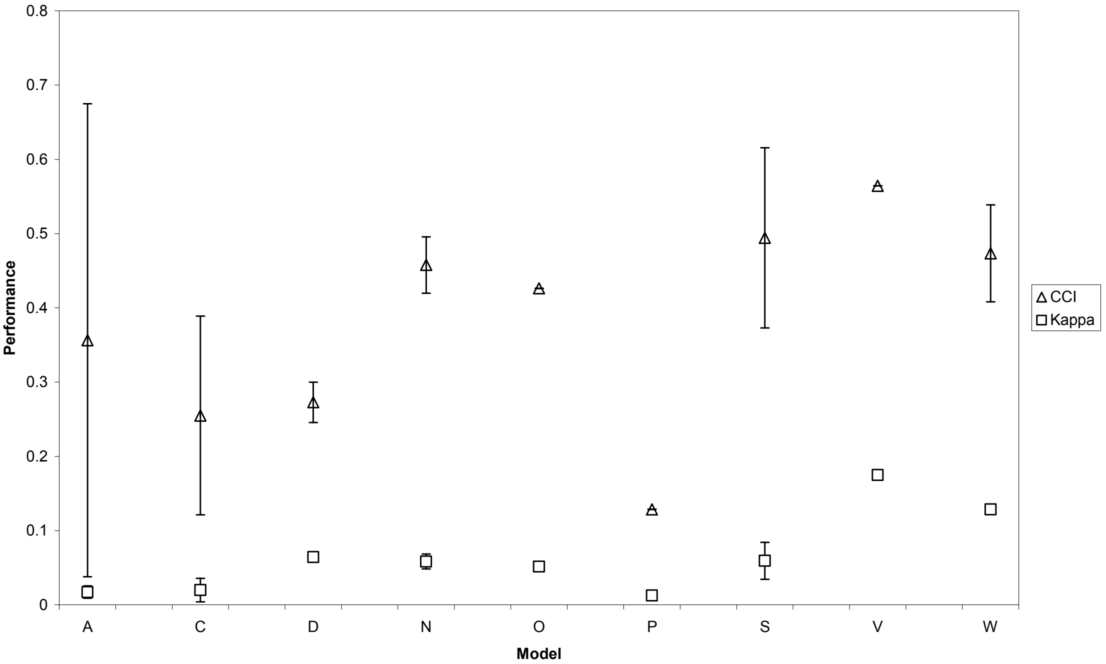


Fig. 5.3. Performance of the 9 data-driven habitat suitability models which incorporated one variable: the ammonium concentration (A), conductivity (C), river depth (D), nitrate concentration (N), dissolved oxygen concentration (O), phosphate concentration (P), the dominant river bed substrate (S), flow velocity (V) and river width (W). Model performance was averaged over the three folds and quantified by the percentage of correctly classified instances (*CCI*) and by *Kappa*.

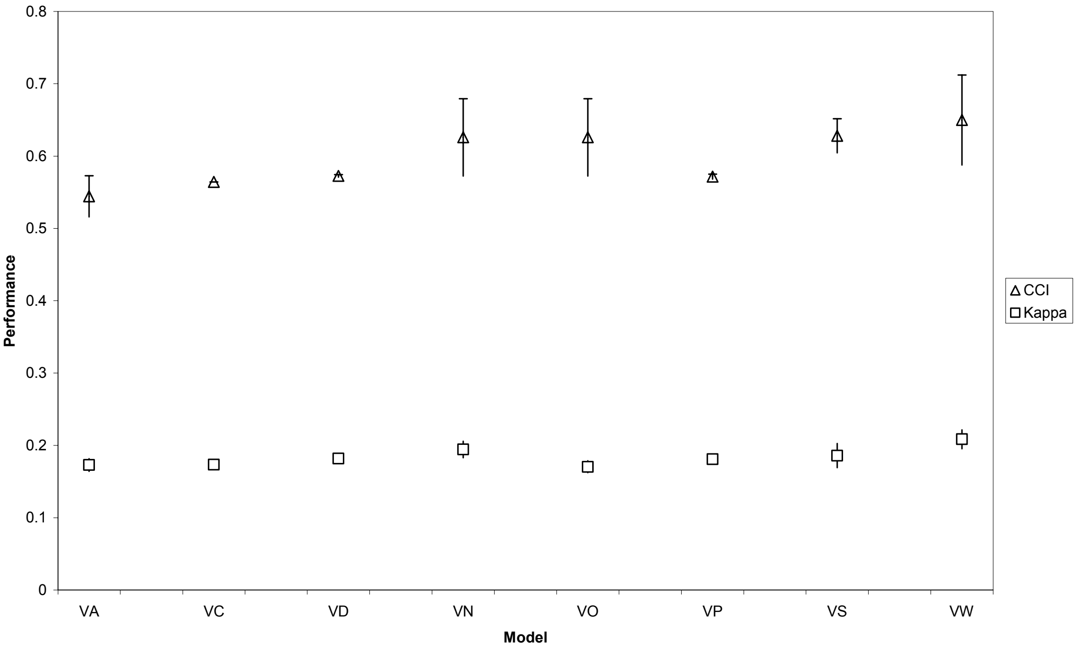


Fig. 5.4. Performance of the 8 data-driven habitat suitability models which incorporated flow velocity (V) and another variable: the ammonium concentration (A), conductivity (C), river depth (D), nitrate concentration (N), dissolved oxygen concentration (O), phosphate concentration (P), the dominant river bed substrate (S) and river width (W). Model performance was averaged over the three folds and quantified by the percentage of correctly classified instances (*CCI*) and by *Kappa*.

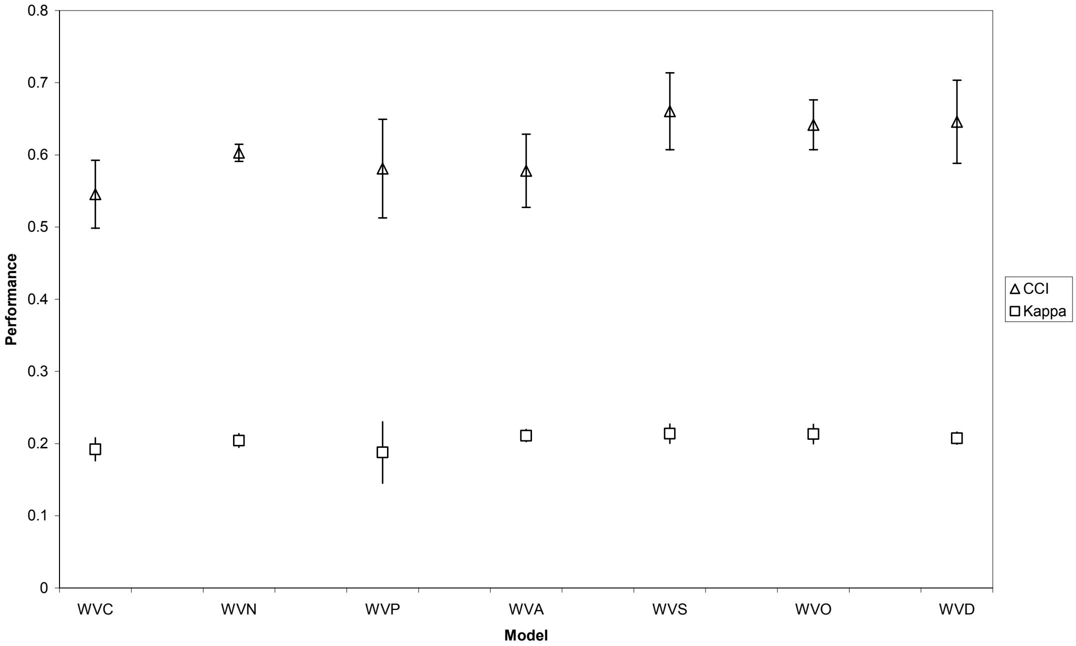


Fig. 5.5. Performance of the 7 data-driven habitat suitability models which incorporated flow velocity (V), river width and another variable: the ammonium concentration (A), conductivity (C), river depth (D), nitrate concentration (N), dissolved oxygen concentration (O), phosphate concentration (P) and the dominant river bed substrate (S). Model performance was averaged over the three folds and quantified by the percentage of correctly classified instances (*CCI*) and by *Kappa*.

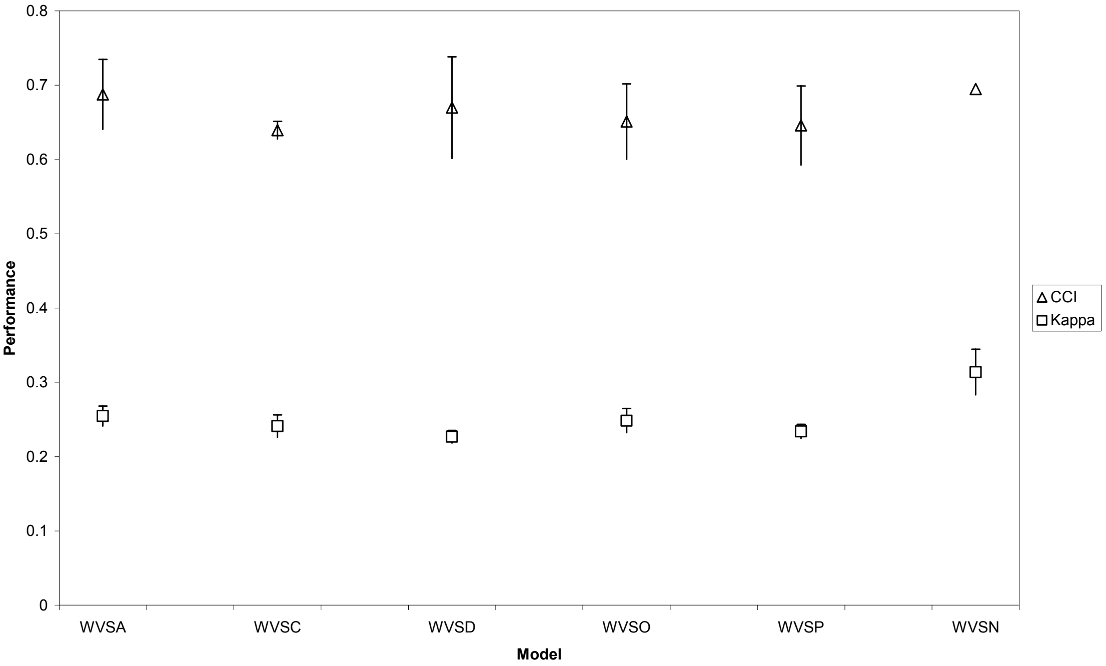


Fig. 5.6. Performance of the 6 data-driven habitat suitability models which incorporated flow velocity (V), river width, the dominant river bed substrate (S) and another variable: the ammonium concentration (A), conductivity (C), river depth (D), nitrate concentration (N), dissolved oxygen concentration (O) and the phosphate concentration (P). Model performance was averaged over the three folds and quantified by the percentage of correctly classified instances (*CCI*) and by *Kappa*.

The ecological relevance of the different models was analysed by plotting the cumulative predicted habitat suitability classes of each model for each input variable. Comparison of these plots with the cumulative observed habitat suitability classes reveals the extent to which the model is over- or underestimating the observations (Figs. 5.7-5.9). Moreover, comparison of the shape of these plots may provide an indication of the ecological relevance of the model. Specifically, a model is assumed to be ecologically more relevant if the shape of its resulting cumulative prediction curve resembles the shape of the cumulative observed habitat suitability and if it shows no underprediction of this habitat suitability. Indeed, an overpredicting model is ecologically irrelevant, whereas overprediction of the observations may not necessarily imply a model error, which is discussed into detail in Chapter 6. Models that show no underprediction and that result in a cumulative plot which resembles the cumulative observations may systematically overpredict the observations. Specifically, if the shape of the plots is similar, the magnitude of the overprediction error is similar over the whole range of the considered input variable. Consequently, the more similar the shape of the cumulative prediction plot is to the shape of the cumulative observation plot, the more ecologically relevant the corresponding model is.

For all variables, the expert knowledge-based model significantly underestimated the observations, whereas the model incorporating flow velocity only (model V) was overestimating the observations the most. Incorporation of other variables (river width; model VW and river bed substrate; model VWS) increased model performance, but further extension of the model by including the nitrate concentration (model WVSN) did not lead to a significantly better model (Figs. 5.7-5.9). This was also reflected in the AIC values (see Section 4.5) of the different models (Table 5.3).

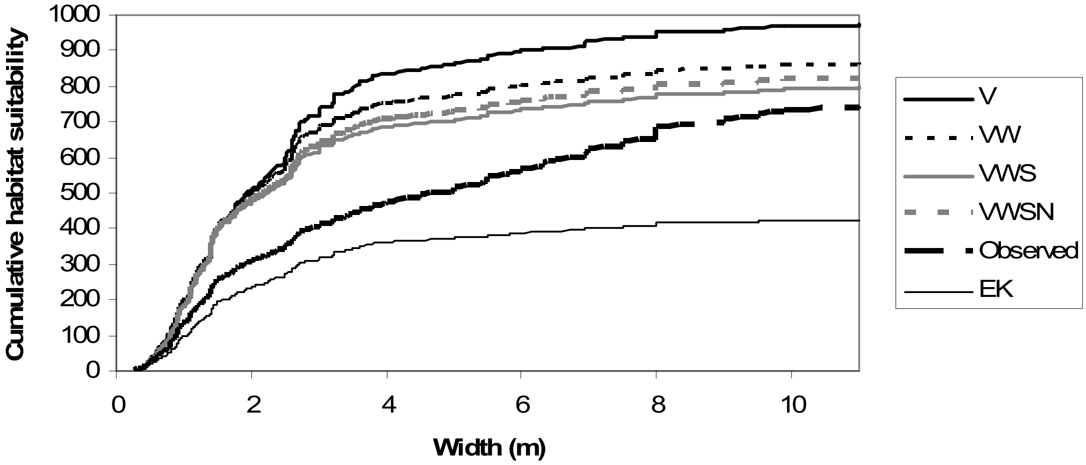


Fig. 5.7. Cumulative predicted habitat suitability classes of the models selected in the model development process compared to the cumulative observed habitat suitability classes as a function of the river width. One out of the five selected models was based on expert knowledge (EK), whereas the other four models were data-driven and either incorporating flow velocity only (model V), flow velocity and river width (model VW), flow velocity, river width and the dominant river bed substrate (model WVS) or flow velocity, river width, the dominant river bed substrate and the nitrate concentration (model WVSN). The 4 habitat suitability classes of the expert knowledge-based model were transformed to the 3 fuzzy sets of the data-driven model output to allow visualisation on the same figure.

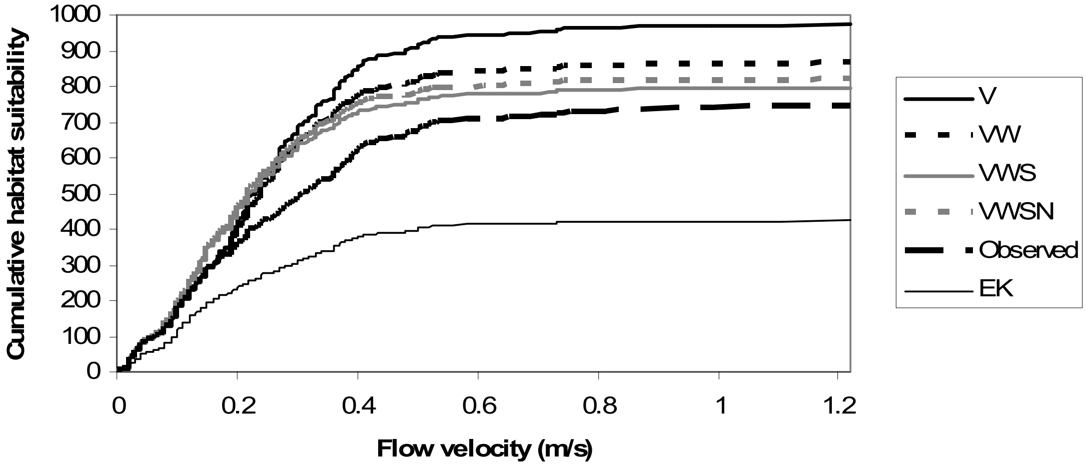


Fig. 5.8. Cumulative predicted habitat suitability classes of the models selected in the model development process compared to the cumulative observed habitat suitability classes as a function of the flow velocity. One out of the five selected models was based on expert knowledge (EK), whereas the other four models were data-driven and either incorporating flow velocity only (model V), flow velocity and river width (model VW), flow velocity, river width and the dominant river bed substrate (model WVS) or flow velocity, river width, the dominant river bed substrate and the nitrate concentration (model WVSN). The 4 habitat suitability classes of the expert knowledge-based model were transformed to the 3 fuzzy sets of the data-driven model output to allow visualisation on the same figure.

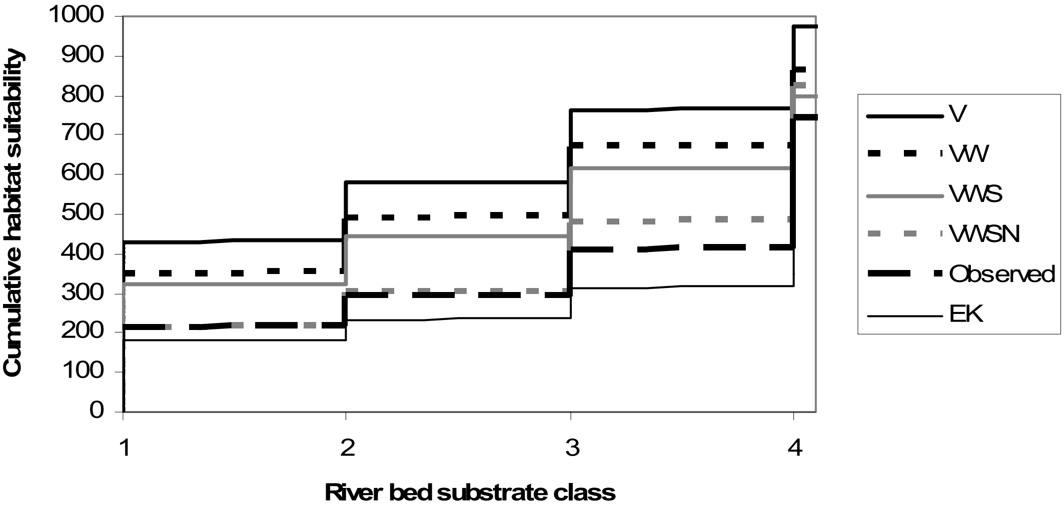


Fig. 5.9. Cumulative predicted habitat suitability classes of the models selected in the model development process compared to the cumulative observed habitat suitability classes as a function of the dominant river bed substrate class. One out of the five selected models was based on expert knowledge (EK), whereas the other four models were data-driven and either incorporating flow velocity only (model V), flow velocity and river width (model VW), flow velocity, river width and the dominant river bed substrate (model WVS) or flow velocity, river width, the dominant river bed substrate and the nitrate concentration (model WVSN). The values of substrate class 4 have been expanded to 4.1 to improve visualisation. The 4 habitat suitability classes of the expert knowledge-based model were transformed to the 3 fuzzy sets of the data-driven model output to allow visualisation on the same figure.

Both model VWS and model VWSN performed similarly for the variables river width, flow velocity and nitrate concentration. Specifically, the highest similarity of both models occurred in the range of river width, flow velocity and nitrate concentration which contained most of the data points (Figs. 5.7-5.9) and the difference between the models only increased substantially at extremely high values of these three variables. However, model WVSN classified more habitats with a finer river bed substrate (dominant substrate class 3 and 4; sand and clay/loam respectively) as suitable for mayflies, whereas model WVS classified more habitats with a coarser substrate (dominant substrate 1 and 2; pebbles and gravel respectively) as suitable.

Table 5.3. Degrees of freedom and AIC values of the four optimal models obtained after addition of variables to the initial one-variable model.

Model	V	VW	VWS	VWSN
Number of degrees of freedom	3	12	48	144
AIC	0.998	-60.879	-68.736	96.132

5.4 Discussion

The presented results suggest that data-driven models may perform better than expert knowledge-based models. This lack of consistency of the expert knowledge, which is confirmed by several authors (Acreman and Dunbar, 2004; Adriaenssens, 2004; Wiens and Graham, 2005; Hernandez et al., 2006; Randin et al., 2006; Fitzpatrick et al., 2007; Strauss and Biedermann, 2007), may hamper the application of expert models. The model development approach in this chapter aims to further contribute to a data-driven knowledge acquisition. Specifically, it showed that even if expert knowledge is available, data analysis could support modellers to generate more reliable habitat suitability models.

Despite a substantial increase in model performance after addition of each variable, the last added variable always was the least important variable in the selected models. Consequently, the model development process only followed a limited part of Algorithm 4.3. However, the other loops of the algorithm ensure that the algorithm reveals the same end point if different starting points are applied. This was tested by starting the algorithm from the best expert knowledge-based model, consisting of flow velocity, river width and nitrate concentration (WVN). The algorithm selected WVNS as the model with the highest marginal increase of performance, and in this model the nitrogen concentration appeared to contribute the least to the total model performance. This variable was thus removed from the model and the algorithm was restarted with the resulting model (WVS), which led to the aforementioned results (Figs. 5.7-5.9).

Although the presented model selection algorithm showed to generate consistent results, care should be taken in selecting an appropriate model. Incorporating more variables into a model will most likely increase model performance, but not necessarily model reliability. This is illustrated by the difference between model VWS and model VWSN. The cumulative model predictions tend to approach the observations increasingly after incorporation of a new variable. However, incorporation of a fourth variable results in a model which generates better *CCI* and *Kappa* values, but the ecological relevance of this model has not necessarily increased. Specifically, the gap between the cumulative model predictions and the observations increased for the river width, flow velocity and the nitrogen concentration, while for the dominating substrate, model WVSN tended to overfit the observations.

The predicted preference of mayfly larvae for higher flow velocities and narrower streams is in line with previous research results (Bengtsson, 1988; Fjellheim, 1996; Mobes-Hansen and Waringer, 1998; Wagner et al., 2000). Moreover, both nitrate concentration and river width may indicate general river degradation because more anthropogenicaly disturbed streams often show higher nitrate concentrations and are further downstream of the stream source and thus wider (Vannote et al., 1980). Since *Baetis* larvae are known to prefer pristine to moderately polluted streams (Bengtsson, 1988), the model predictions for river width and nitrate concentration appear to be ecologically relevant.

However, further analysis of the model predictions is needed based on the dominant substrate, because these predictions are less similar among different models. Several authors showed that *Baetis rhodani* larvae prefer coarse substrate over finer substrates (Bengtsson, 1988; Fjellheim, 1996; Mobes-Hansen and Waringer, 1998; Wagner et al., 2000). These results are also confirmed by Wood et al. (2005), who revealed that these mayfly larvae are unable to excavate themselves from sedimented material and thus avoid dynamical substrates. Another explanation for the coarse substrate selection of mayfly larvae may be antipredatory behaviour (Tikkanen et al., 1996; Dahl, 1998; Huhta et al., 1999; Kelly et al., 2002). Tikkanen et al. (2000) also reported that a majority of *Baetis* larvae chooses dark substrates over light-colored ones in both laboratory and natural streams, although they could not attribute these results to antipredatory behaviour.

Given the general agreement on the preference for coarse substrates of mayfly larvae, the predictions of model WVS may be ecologically more relevant than those of model WVSN. This leads to a second important issue for the selection of an appropriate ecological model, which is the distinction between over- and underestimation of the observations. This type of model error describes a consistent under- or overestimation of the species prevalence (Pearce and Ferrier, 2000) and is discussed into detail in Chapter 2 and in the next chapters. Given the considerations on over- and underprediction of the observations, model WVS may be ecologically more relevant than model WVSN. This is in line with the characteristics of the Zwalm river basin, which is one of the least polluted river basins in Flanders and shows a good to moderate chemical water quality. Consequently, the habitat suitability of *Baetis* larvae in this river basin may be quantified reliably by physical habitat variables. Moreover, *Baetis* larvae also occur in rivers with moderate chemical water quality, which supports the aforementioned results. Since in the studied area, wider rivers generally show more anthropogenic disturbance, river width may also roughly reflect chemical water quality. Consequently, the effect of the nitrate concentration may already be partly incorporated in the river width variable, which might explain the relatively low importance of the nitrate concentration in the habitat model.

For some variables, such as the dominant river bed substrate, model WVSN appears to overfit the training data, although this overfitting is not in line with the aforementioned habitat preferences of *Baetis* larvae. Model WVSN is prone to overfitting due to its relatively high number of model parameters (144 parameters for 319 data points). This was also reflected by the AIC values of the different models, which quantify the balance between model performance and complexity (Akaike, 1974). Although this measure assigns an arbitrary weight to both model characteristics, it provides an indication of how likely a model will generate overfitted predictions. Moreover, a higher number of model parameters compared to the number of sampling points increases the likelihood that model training is stopped at a local optimum, and that the global optimum is never reached. Finally, the development of reliable but more complex models often requires more data, more monitored variables and the incorporation of a wider environmental range in the training data, which results in a substantial increase in costs and efforts.

This chapter shows that data-driven model development may complement the expert knowledge approach which is often used in habitat suitability modelling of aquatic species. The presented method was applied on a mayfly species in a specific river basin, but it could be easily applied to any species, river system and at different spatial and temporal scales. The flexibility of the approach allows modellers to select habitat variables and to define their classification in terms of fuzzy sets specific to the study site, the research objectives or the data availability. The approach in this chapter allows quantitative description of expert knowledge, and thus enables comparison of the data-driven habitat suitability predictions with available expert knowledge. Consequently model users could choose the most appropriate model based on the quality of the available data and expert knowledge. Data-driven rules may be more reliable if a substantial number of high quality samples is available, whereas expert knowledge based rules may reliably predict habitat suitability in study areas which are underrepresented in the available dataset. To enable visualisation of the predicted habitat suitability, the final fuzzy rules could be implemented in visualisation tools such as MesoCASiMiR module of the CASiMiR modelling system (Jorde et al., 2000; Schneider, 2001; Mouton et al., 2006; Mouton et al., 2007). As such, the presented method may provide a valuable tool for river managers and stakeholders to select different restoration options and to implement their management strategies.

CHAPTER 6

Theoretical analysis of performance criteria

6.1 Introduction

The previous chapter suggested that data-driven model development may complement the expert knowledge approach which is often used in habitat suitability modelling of aquatic species. To develop these data-driven models, however, a sound comparison of model performance is needed to allow reliable model training and evaluation of the final model. A crucial step in the model comparison procedure is the assessment of the model performance (Fielding and Bell, 1997; Manel et al., 1999a; Manel et al., 1999b; McPherson and Jetz, 2007). Most authors refer to this step as model evaluation (Boyce et al., 2002; Anderson et al., 2003; Barry and Elith, 2006; Guisan et al., 2007), hereby situating the model evaluation procedure at the end of the overall modelling process. However, model performance is also assessed during the model development process to compare trained models and select the best performing models (Hastie et al., 2001). To avoid misleading terminology in this chapter, the term *model evaluation* will refer to performance assessment of the final model, whereas *training* performance assessment will refer to model performance assessment during model training.

The key component of model performance assessment is the criterion applied to quantify model performance. Since Fielding and Bell (1997) reviewed the performance criteria most commonly applied in conservation presence/absence methods, the performance assessment of models developed from presence-absence data has been a recurrent focus (Pearce and Ferrier, 2000; Manel et al., 2001; Nielsen et al., 2005; Vaughan and Ormerod, 2005; Allouche et al., 2006). Estimates of predictive performance have been widely applied to assess final model quality, especially in papers that compare the relative performance of different methods (Hirzel et al., 2001; Elith et al., 2006; Heikkinen et al., 2007; Meynard and Quinn, 2007; Peterson et al., 2007). Most literature on training performance assessment focuses on measures such as Akaike's information criterion, which facilitates variable selection by estimating the relative error rates for alternative models built using the same dataset (Hastie et al., 2001). These papers describe the selection of fitted models, whereas the focus of this chapter is on the model performance assessment during model training. An evaluation of a sample of species distribution modelling literature between 1998 and 2007 (*n* = 385), indicated that 67 % of the modellers use data for model training or evaluation. Although the model fitting procedure is mostly (99 %) based on optimisation of the predictive accuracy of the model, a wider range of performance criteria has been used for model evaluation (Table 6.1). Manel et al. (2001) reviewed a sample (*n* = 87) of published ecological literature between 1989 and 1999, which revealed that many users of presence-absence models made no evaluation at all, even in leading ecological journals. A more recent review in this chapter shows that not only the number of modellers applying model evaluation increased substantially in recent years, but also that more than 82 % of these modellers applied at least two different performance criteria for model evaluation (Table 6.1).

This chapter aims to review the performance criteria most commonly applied in presence-absence modelling and analyse their function in both the model training and the model evaluation process. First, an up-to-date general framework for error quantification in habitat models is provided. Then, this framework is related to the role of performance criteria in both model evaluation and model training. The role of performance criteria in model evaluation is reviewed and some empirical insights are given which are important for both model training and evaluation. Furthermore, the role of performance criteria in model training is analysed theoretically and some pitfalls are identified. Although the focus of this chapter is on presence-absence models, some concepts and problems which also occur across ecological modelling more generally, will be discussed. Finally, recommendations are given on the application of different performance criteria on model evaluation and training.

6.2 A framework for evaluating model performance

Model performance assessment, based on the quantification of error and uncertainty in models, has increasingly received attention during the last decade (Vaughan and Ormerod, 2005; Barry and Elith, 2006). Different authors focused on either the theoretical source of modelling errors (Guisan and Zimmermann, 2000; Austin, 2007), the performance criteria for model evaluation (Fielding and Bell, 1997; Pearce and Ferrier, 2000; Manel et al., 2001) or both (Vaughan and Ormerod, 2005; Barry and Elith, 2006). This chapter represents a further contribution to the latter category. Murphy and Winkler (1987; 1992) have developed a framework for assessing predictive performance of models that explicitly links model reliability and discrimination ability. By partitioning out the different error components, this framework may provide more useful information for further improving models (Guisan et al., 2006). The framework was reviewed in an ecological context by Pearce and Ferrier (2000) and has been applied in many other studies on predictive performance of species distribution models (Vaughan and Ormerod, 2005).

The framework, summarised in Fig. 6.1, allows prediction error to be partitioned between a number of sources. These model error sources are either based on the observations or on the model predictions (Pearce and Ferrier, 2000). Model discriminatory ability and prevalence belong to the first category, while model calibration and refinement are included in the latter.

Refinement relates to the range of predictions produced by the model for a given set of sites. Predictions from most distribution models are continuous variables, even if they are fitted to presence-absence data. Good refinement indicates that predictions cover the full probability range, with predicted values near both zero and one (Pearce and Ferrier, 2000).

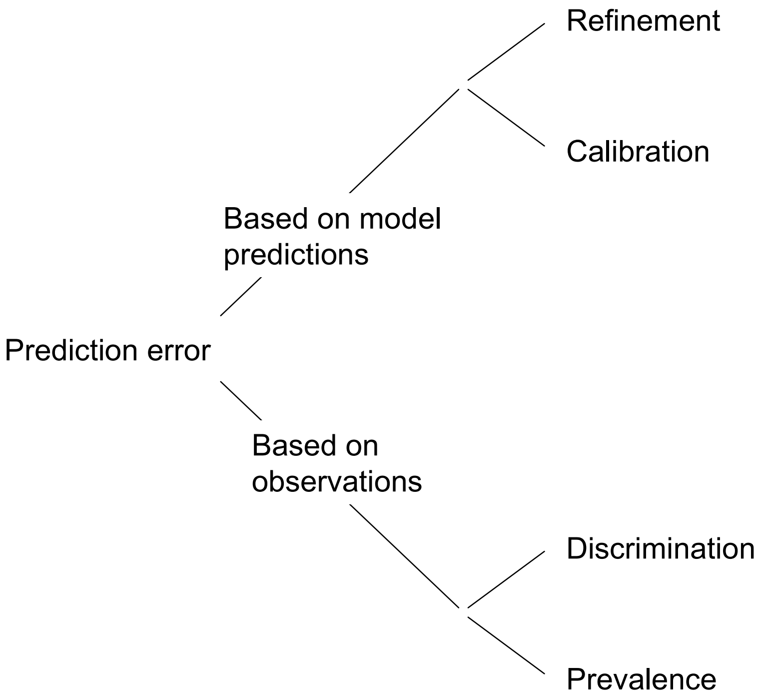


Fig. 6.1. A framework for evaluating model performance (adjusted from Pearce and Ferrier (2000)).

Calibration refers to the numerical accuracy of the predictions (Harrell et al., 1996), which reflects the level of agreement between actual observations and model predictions (Pearce and Ferrier, 2000). A model shows good calibration if, for instance, sites given predicted probability values of 0.8 have an 80 % chance of being occupied and whether this is twice as likely for sites given predictive probability values of 0.4 (Vaughan and Ormerod, 2005). The source of poor calibration can be partitioned into three components: bias, spread and unexplained error.

Bias describes a consistent under- or overestimation of the species prevalence and is also referred to as omission or commission error (Rondinini et al., 2006) or as false-negative or false-positive error (Loiselle et al., 2003), respectively. Several authors have addressed this issue in ecological modelling and emphasized the significant negative correlation between both errors (Fielding and Bell, 1997; Anderson et al., 2003; Loiselle et al., 2003; Rondinini et al., 2006; Fitzpatrick et al., 2007). Consequently, conservationists should choose between models that minimise one or the other error (Loiselle et al., 2003; Wilson et al., 2005; Rondinini et al., 2006). Knowledge on the data which have been used to develop the model may substantially influence this choice because these data can inherently contain either more commission or omission errors, or both. Data derived from geographic ranges, for example, contain large commission errors since these ranges assume homogeneous species distributions. Although point locality data obtained from museum collections generally contain large omission errors, point locality data from field observations can reduce commission errors. False-positive observations are unlikely because these errors can only occur if samples are contaminated (Rondinini et al., 2006). Sample contamination could occur more easily for smaller species, but in most cases this error could be avoided by applying good sampling practice. Sampling inefficiency could lead to false-negative observations, which falsely suggest that a suitable habitat is not suitable for a species of interest. If not adequately dealt with, false-negative observations are likely to affect both variable selection and coefficient estimation in models (MacKenzie et al., 2002; MacKenzie et al., 2003; Tyre et al., 2003; Barry and Elith, 2006).

Boyce et al. (2002) emphasize that mobile animals may not be using the entire habitat at one time and that modelling their habitat thus requires an appropriate modelling technique. Specifically, false-positive prediction of the observed situation by a model may not necessarily imply a model error. Several factors suggest these false-positive predictions are inherent to the classification of ecological data. Monitoring inefficiency is an explanation (MacKenzie et al., 2003), but the occurrence of aquatic species may also depend on other variables than those included in the habitat suitability model (Gönczi, 1989). For example, if a fish species is absent due to the poor water quality at a site with excellent physical habitat, a good physical habitat model will predict that this fish is present at this site. Several authors suggest that species might be temporarily absent from the site (Lütolf et al., 2006) or the site might not yet be colonized (Hirzel et al., 2001). This problem could be solved by including a temporal variable such as seasonality or the intensity of past disturbance in the model. Further research should provide more insight into this gap between expert knowledge and the observed occurrence of aquatic species and into the assessment of the direct gradients which determine species occurrence.

In contrast to the false-positive predictions, only few hypotheses can explain false-negative predictions of habitat suitability models. If an aquatic species is present at a specific site, models predicting this species to be absent at this site are always false, assuming that the biological data contain no errors. Consequently, overpredicting models may be more realistic than models which underestimate the actual occurrence of a species. However, the acceptance of false-positive predictions should be limited, since models predicting a species to be present at all sites are ecologically irrelevant, assuming that the data cover both suitable and unsuitable habitats (Manel et al., 2001).

Loiselle et al. (2003) indicate that conservation decision makers must decide which error is most unlikely. They argue that commission errors may lead to failure to conserve a species because sites selected as reserves do not contain the target taxa. Omission errors may fail to select sites that are of importance to the overall survival of a species, such as those that represent viable populations (Loiselle et al., 2003). They suggest to select species distribution models that minimise false-positive errors to avoid the selection of reserve areas that do not actually contain the target taxa. However, this implies that models with a higher false-negative error are selected given the significant negative correlation between omission and commission errors. As mentioned before, such models may fail to identify sites that are of importance to the overall survival of the species, which contrast markedly with the precautionary principle that is often applied in ecological decision making (Wilson et al., 2005). Hence, selection of models minimising underprediction may be a more prudent solution. In a modelling study of the potential distribution of two rodents, (Anderson et al., 2003) reported that the best models were consistently found at low levels of omission and moderate-to-high commission values. Applying presence-only data, they suggest to arbitrarily set an acceptable level of omission error. In contrast, if the purpose of the species distribution model is to select suitable regions for reintroduction of a locally extinct species, commission errors should be avoided because these would lead to reintroduction in areas which are not suitable for the species of interest (Pearce and Ferrier, 2000).

The second component of calibration, spread, indicates that the model is systematically overestimating and underestimating observed values, depending on whether the predicted probabilities are higher or lower than 0.5. Predictions greater than 0.5 are underestimating the occurrence of the species and predictions less than 0.5 are overestimating the occurrence of the species, or vice versa. If a model has a significant bias or spread error, a predicted value of 0.5 will not relate to an observed value of 0.5, but to a higher or lower observed value. Therefore, choosing a threshold probability without any information on bias and spread will greatly reduce the model confidence. Moreover, the presence of spread error and lack of model refinement suggest that important explanatory variables are missing from the model (Pearce and Ferrier, 2000).

Unexplained error, the last component of calibration, describes variation not accounted for by the bias and the spread of a model. Some of this variation may arise because variables that may influence species occurrence were not included in the model or due to random variation, which is increased by error in measurement of species presence and environmental variables (Pearce and Ferrier, 2000).

The second component of model error, which is based on the observations, consists of the discrimination of the model and of the prevalence. Discrimination relates to the ability of a model to distinguish occupied from unoccupied sites (Harrell et al., 1996; Pearce and Ferrier, 2000). It is the probability that a model will assign higher predicted values to occupied sites compared with those assigned to unoccupied sites. The prevalence indicates how often a species is present at a sampled set of sites and therefore provides information on the probability of a species being observed as present at a randomly selected site (Murphy and Winkler, 1987; Pearce and Ferrier, 2000).

6.3 Performance criteria in presence-absence modelling

6.3.1 Selected sources of model error

Although model error sources based on the model predictions are related to error sources based on observations (Pearce and Ferrier, 2000), these two error types reflect different aspects of the predictive performance of the model. The relative importance of both error types depends on the purpose of the model (Pearce and Ferrier, 2000; Vaughan and Ormerod, 2005). Pearce and Ferrier (2000) emphasize that a model which has good calibration and refinement must also have good discrimination at a constant prevalence, but acknowledge that the converse is not necessarily true. This chapter will focus on the model discrimination as a source of model error, because this is the fundamental component of prediction accuracy (Vaughan and Ormerod, 2005). Although it is necessary to have at least a moderate level of refinement in order to be able to examine model performance further, few authors treat it as a substantial source of model error (Vaughan and Ormerod, 2005) because adjustment of the model structure can easily avoid a low level of refinement. Model calibration is only required if predictions are used in a quantitative matter, for instance to generate spatial maps of the probability of species occurrence (Johnson et al., 2004). Often, a model may only be required to rank sites according to their relative probability of being occupied: treating its predictions as being ordinal, rather than quantitative. In such instances, a test of discrimination may be deemed sufficient (Vaughan and Ormerod, 2005).

This chapter distinguishes the prevalence of the training and evaluation dataset from the three aforementioned sources of model error because it is the only model error source which cannot be reduced by adjusting the model structure (Pearce and Ferrier, 2000). The dependency of different performance criteria values on prevalence has been widely acknowledged (Manel et al., 2001; McPherson et al., 2004; Luoto et al., 2005; Allouche et al., 2006; Bulluck et al., 2006). This dependency has a significant impact on the ability to compare accuracy meaningfully between the same model in different applications (with different species prevalence) or between models developed for different species or with different training or evaluation data (Vaughan and Ormerod, 2005). Many authors suggested that the prevalence should approximate 0.5 in order for the predictive performance of a model to be examined (Pearce and Ferrier, 2000; Liu et al., 2005; Maggini et al., 2006). However, others suggested that performance criteria should be either largely independent of, or possibly corrected for species prevalence (Miller et al., 1991; Manel et al., 2001; Allouche et al., 2006).

6.3.2 Performance criteria

The key component of the model training and validation procedures is the performance criterion which evaluates the model performance. Performance criteria can deal with either continuous or discrete model outputs, or with both. All performance criteria which were developed to evaluate discrete model predictions can also handle continuous predictions because these can be discretised by using threshold values. Generally, suitable sites, where species are supposed to be present, are separated from unsuitable sites by one single habitat suitability threshold. The many approaches to determining thresholds fall into two categories: subjective and objective. Several authors apply the first category, for instance by taking 0.5 as the threshold (Manel et al., 1999a; Manel et al., 2001; Bailey et al., 2002; Luck, 2002; Stockwell and Peterson, 2002; Woolf et al., 2002; Dedecker et al., 2005). More objective approaches choose thresholds to maximise the agreement between observed and predicted distributions. Liu et al. (2005) compared 12 different objective approaches to determining thresholds and investigated the effect of different prevalence for training and evaluation data. They recommend using either the prevalence approach, the average probability approach, or a sensitivity and specificity-based approach (Liu et al., 2005). The prevalence approach (Cramer, 2003) takes the prevalence of the training data as a threshold, whereas the average probability approach (Cramer, 2003) takes the average predicted probability of the training data as a threshold. Sensitivity and specificity-combined approaches either maximise the sum of both (Cantor et al., 1999; Manel et al., 2001), minimise the absolute value of the difference between *Sn* and *Sp* (Cantor et al., 1999), or derive the threshold from the Receiver-Operator Characteristic (ROC) plot (Cantor et al., 1999). The commonly applied *Kappa* maximisation approach, which chooses the threshold value that maximises *Kappa* (Guisan et al., 1998; Hirzel et al., 2006), appeared to be less appropriate, while the fixed threshold approach turned out to be the worst method (Liu et al., 2005). In a smaller comparison of four threshold selection methods, Jiménez-Valverde and Lobo (2007) found similar results.

Despite the aforementioned methods, the choice of an appropriate threshold often remains difficult and arbitrary (Fielding and Bell, 1997; Manel et al., 1999b; Manel et al., 2001; Liu et al., 2005) and depends on the calibration of the model (Pearce and Ferrier, 2000; Vaughan and Ormerod, 2005). Moreover, selection of a threshold often depends on the conservationist's preferences and can significantly affect reserve selection for conservation planning (Liu et al., 2005; Wilson et al., 2005). Therefore, some criteria, such as the average deviation (*AD*; Van Broekhoven et al. (2007)), do not require this arbitrary threshold to process continuous data. The receiver operator characteristic (ROC; Miller et al., 1991; Fielding and Bell, 1997) approach is another alternative method for assessing the accuracy of probabilistic output models (Fig. 6.16). ROC curves are constructed by using all possible thresholds to classify the model outputs into confusion matrices, obtaining the sensitivity and specificity for each matrix, and then plotting sensitivity against the corresponding proportion of false-positives (equal to 1-specificity). The area under the ROC curve (*AUC*) is often used as a single threshold-independent measure for model performance (Manel et al., 2001; Thuiller et al., 2003; Brotons et al., 2004; McPherson et al., 2004; Thuiller et al., 2005). *AUC* was shown to be independent of prevalence (Manel et al., 2001; McPherson et al., 2004) and is an effective measure of discriminatory ability for probabilistic models (Vaughan and Ormerod, 2005). Consequently, *AUC* is considered to be the current best practice for assessing model success for presence/absence data (Pearce and Ferrier, 2000; Thuiller et al., 2003; Rushton et al., 2004; Austin, 2007). However, the *AUC* approach cannot be applied to dichotomous presence-absence model outputs (Allouche et al., 2006). Moreover, models with the same or very similar *AUC* values may predict very different patterns of distribution (see further, Fig. 6.16). Finally, Maggini et al. (2006) found that the *AUC* is systematically lower at extreme prevalence values (prevalence < 0.05 or > 0.70). The *AUC* appears to be independent of prevalence only in its middle range (Maggini et al., 2006; McPherson and Jetz, 2007). Reliance on *AUC* as a sufficient test of model success needs to be re-examined (Termansen et al., 2006; Austin, 2007).

The most popular measure for the accuracy of presence-absence predictions is Cohen's *Kappa* (Manel et al., 2001; Loiselle et al., 2003; Petit et al., 2003; Berg et al., 2004; Parra et al., 2004; Pearson et al., 2004; Rouget et al., 2004; Segurado and Araújo, 2004; Allouche et al., 2006). This measure allows an assessment of the extent to which models correctly predict occurrence at rates that are better than chance expectation (Fielding and Bell, 1997). However, several authors argued that *Kappa* may be less appropriate for model evaluation due to its dependence on the prevalence of the training data from which the model was developed (Fielding and Bell, 1997; Manel et al., 2001; Allouche et al., 2006). Three performance criteria were proposed to avoid this problem because these criteria were assumed to be independent of prevalence (Manel et al., 2001; Allouche et al., 2006): the *NMI* (Forbes, 1995; Fielding and Bell, 1997), *TSS* (Allouche et al., 2006) and *AUC* (Fielding and Bell, 1997; Manel et al., 2001).

Several authors suggested some desirable properties of accurracy statistics for the assessement of species distribution model performance (Forbes, 1995; Fielding and Bell, 1997; Vaughan and Ormerod, 2005). According to Vaughan and Ormerod (2005), the most important property is generality, which is defined as the ability to compare accuracy meaningfully between the same model in different applications or between models developed for different species or with different training and test data. This property has been defined from an evaluation perspective: a suitable performance criterion should be independent of the prevalence of the data to which the criterion is applied. This generality attempts to avoid that two identical models which are evaluated on two different datasets, would show different model performance. Vaughan and Ormerod (2005) suggest that, if performance criteria values are affected by prevalence, the criteria could also be corrected for this prevalence to assure generality. Another desirable property of performance criteria is the ability to weigh omission and commission errors. It is generally accepted that conservationists should be able to choose between both errors (Loiselle et al., 2003; Wilson et al., 2005; Austin, 2007) and that successful recovery of the true model depends more on the ecological insight of the modellers than on the particular statistical modelling method used (Austin et al., 2006; Austin, 2007). Consequently, suitable performance criteria should allow species distribution modellers to take this issue into account. In the next sections, the suitability of the most frequently applied performance criteria for both the evaluation and training of species distribution models is assessed based on their generality and their distinction between omission and commission errors.

6.3.3 Application of performance criteria for model evaluation

Recent research on species distribution modelling often compares the predictive performance of different habitat suitability models (Manel et al., 1999a; Manel et al., 1999b; Elith et al., 2006; Meynard and Quinn, 2007). This comparison is based on the assessment of the performance of the final model which is obtained after model training. Most authors refer to this step as model evaluation (Boyce et al., 2002; Anderson et al., 2003; Barry and Elith, 2006; Guisan et al., 2007), although the terms model testing and model validation are also being used (Anderson et al., 2003; Vaughan and Ormerod, 2005). Both latter terms are less appropriate to designate model performance assessment because they may overlap with other steps in the modelling process. In neural network applications for instance, model testing is applied to stop the supervised learning procedure and to avoid overfitting of the training data, which occurs when idiosyncrasies in the training set are modelled in addition to the underlying species-environment relationship (Lek and Guégan, 1999). This results in a misleadingly good fit to the data (Geman et al., 1992; Harrell et al., 1996; Gevrey et al., 2003). Model validation implies the quantification of the model performance on an independent dataset. Optimally, this dataset should be completely independent from the data used to train or calibrate the model, e.g. collected on other areas (Fielding and Bell, 1997; Hastie et al., 2001). However, due to time and money constraints, most studies have only one dataset which is split in a training and an evaluation set. Moreover, a truly independent validation set is unrealistic because it is impossible to monitor exactly the same ecological processes at the same environmental conditions at different moments in time. Consequently, models are evaluated by *k*-fold cross-validation to allow assessment of the robustness of the model performance quantification (Vanhouwelingen and Lecessie, 1990; Fielding and Bell, 1997; Hastie et al., 2001). This resampling approach randomly divides the dataset into *k* independent parts, using *k*-1 of them to train the model, and calculating the model performance on the left-out part. This produces *k* estimations of the model performance, allowing assessment of its mean and variance. The number of parts typically varies between 3 and *n*, depending on the number of instances *n* (Hastie et al., 2001; Hirzel et al., 2006). The latter situation is referred to as leave-one-out or jack-knife cross-validation.

In recent years, model evaluation has increasingly received attention in species distribution modelling. The aforementioned sample of ecological literature between 1989 and 1999, which was evaluated by Manel et al. (2001), revealed that only 52 % of the modellers (n = 87) evaluated model performance. A more recent sample of ecological literature on presence-absence or presence-only models (n = 385) not only indicated that the number of papers on species distribution modelling increased significantly, but also that 97 % of the model users evaluates model performance (Table 6.1). Araújo et al. (2005) found similar results for a sample of species-climate envelope models under climate change between 1995 and 2004 (n = 29), of which 93 % was evaluated. Although 82 % of the modellers applied two or more criteria, the percentage of papers applying at least two performance criteria has only increased slowly. Table 6.2 shows that almost all papers use *CCI*, but the application of *Kappa* and *AUC* is increasing, with *AUC* being the most widely applied performance criterion for model evaluation. In recent years, some papers introduced other performance criteria into species distribution modelling, such as the odds ratio, the *TSS* and the *NMI* statistic. Current best practice in species distribution modelling is to apply at least two different performance criteria for model evaluation.

Table 6.1. The number of validation criteria used in the model evaluation process. A sample of 385 papers on species distribution modelling was evaluated, of which 67 % used data for model training or evaluation. The numbers in the table are percentages of this group (n = 257). The papers which applied 0 performance criteria did not evaluate model performance.

	Number of performance criteria applied in model evaluation						Total
	0	1	2	3	4	5	
1998-2002	2	18	61	8	8	3	15
2003-2007	3	15	64	12	4	2	85
Total	3	15	64	11	5	2	100

Table 6.2. The different validation criteria used in the model evaluation process. A sample of 385 papers on species distribution modelling was evaluated, of which 67 % used data for model training or evaluation. The numbers in the table are percentages of this group (n = 257). *CCI* = percentage correctly classified instances; *CCI* only = papers which only applied *CCI* for model evaluation; *Kappa* = Cohen's *Kappa*; *AUC* = Area Under the Curve; *Sn* = Sensitivity; *Sp* = Specificity. The percentages do not cumulate to 100 % because some papers used more than one performance criterion.

	Performance criteria applied in model evaluation							
	No criterion	*CCI*	*CCI* only	*Kappa*	*AUC*	*Sn*	*Sp*	Other
1998-2002	2	98	18	26	53	16	16	1
2003-2007	3	97	15	33	61	8	7	3
Total	3	97	15	32	60	9	8	2

Taxonomic groups most frequently involved in presence-absence models were trees and other angiosperms (31 %), birds (26 %), invertebrates (21 %), mammals (17 %), fish (9 %), reptiles (4 %) and amphibians (2 %). Less frequent applications were from bacteria, plankton and fungi. Model and optimisation techniques included regression techniques such as GLM or GAM (84 %), classification trees (17 %), artificial neural networks (14 %), genetic algorithms (13 %) and other methods (12 %).

It has been suggested that some of the performance criteria depend on the prevalence of the evaluation dataset (Luoto et al., 2005; Vaughan and Ormerod, 2005). Substantial work has been done to assess the relation between *Kappa* and prevalence empirically (Manel et al., 2001; McPherson et al., 2004) and theoretically (Allouche et al., 2006), but the relation between some less frequently used performance criteria and prevalence has received far less attention. This chapter attempts to analyse this relation systematically by calculating the performance measure values from 100000 randomly generated confusion matrices. First, the prevalence of each matrix was generated randomly and then the matrices were constructed based on this prevalence. For each matrix, the value of 6 different performance criteria was calculated and is plotted in Figs. 6.3-6.8. Fig. 6.2 shows the distribution of the different prevalence values among 100 different classes.

The relation between *CCI* and prevalence is shown in Fig. 6.3. If the performance criterion would be independent of prevalence, the density of confusion matrices would be similar over the whole prevalence range. However, selection of the areas (squares of 0.01 x 0.01) in which at least x % of the 100000 confusion matrices is situated, reveals that this density is only similar for all *CCI* values at an intermediate prevalence value (P = 0.5). At extreme prevalence values, the confusion matrices tend to show either excellent or extremely poor predictive accuracy. These findings agree with the results of a species distribution modelling study of 18 bird species in Nevada, USA (Bulluck et al., 2006). For *CCI* values greater than 0.5, they found a similar relation between the *CCI* and the prevalences of both the evaluation and training sets.

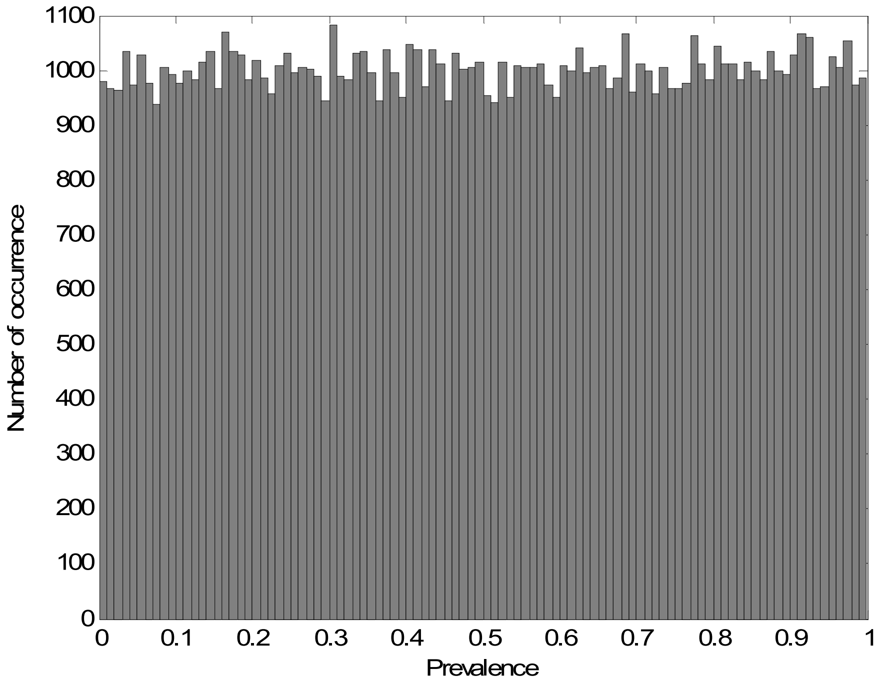


Fig. 6.2. Distribution of the randomly simulated prevalence values over 100 prevalence classes. A perfectly random process should result in a perfectly uniform distribution.

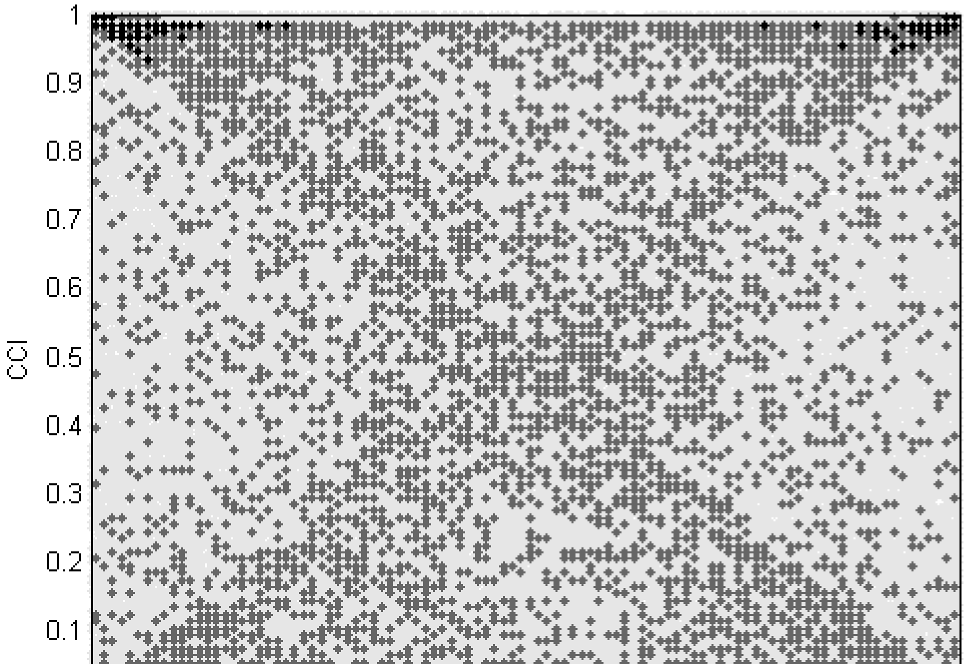


Fig. 6.3. Relation between *CCI* and prevalence, obtained after simulation of 100000 confusion matrices with random prevalence. To reveal patterns in this plot, areas in which the density of confusion matrices exceeded a fixed threshold *x* were selected. The background of this figure is white. Light grey dots indicate all confusion matrices, whereas dark gray and black dots indicate the centre of a 0.01 x 0.01 square in which at least 1 % and 2.5 % of the confusion matrices is situated, respectively.

The same approach confirms that *Sn* also depends on prevalence (Fig. 6.4). At lower prevalence values, more differentiated *Sn* values occur, whereas at higher prevalences, *Sn* tends to be either minimal or maximal. Similar conclusions were found for *Sp*, which showed the same graph mirrored about the x = 0.5 axis. These findings agree with previously reported relations between prevalence and *Sn* or *Sp* (Allouche et al., 2006). Bulluck et al. (2006) found a significant positive quadratic and linear relationship between sensitivity and the respective prevalences of the evaluation data sets and the training sets. Similar results for the training set were obtained by McPherson et al. (2004) while modelling 32 bird species endemic to South Africa, Lesotho and Swaziland. These relations could be a subset of the relation between prevalence and *Sn* which is shown in Fig. 6.4. However, generalisation of these empirical relations may be less appropriate because the prevalence in the study of Bulluck et al. (2006) and McPherson et al. (2004) is not distributed uniformly over the different cases, which contrasts with the prevalence distribution applied in this section (Fig. 6.2). As will become clear, this lack of uniformity in the prevalence distribution may significantly affect the relations which are found in studies analysing the relation between performance criteria and prevalence.

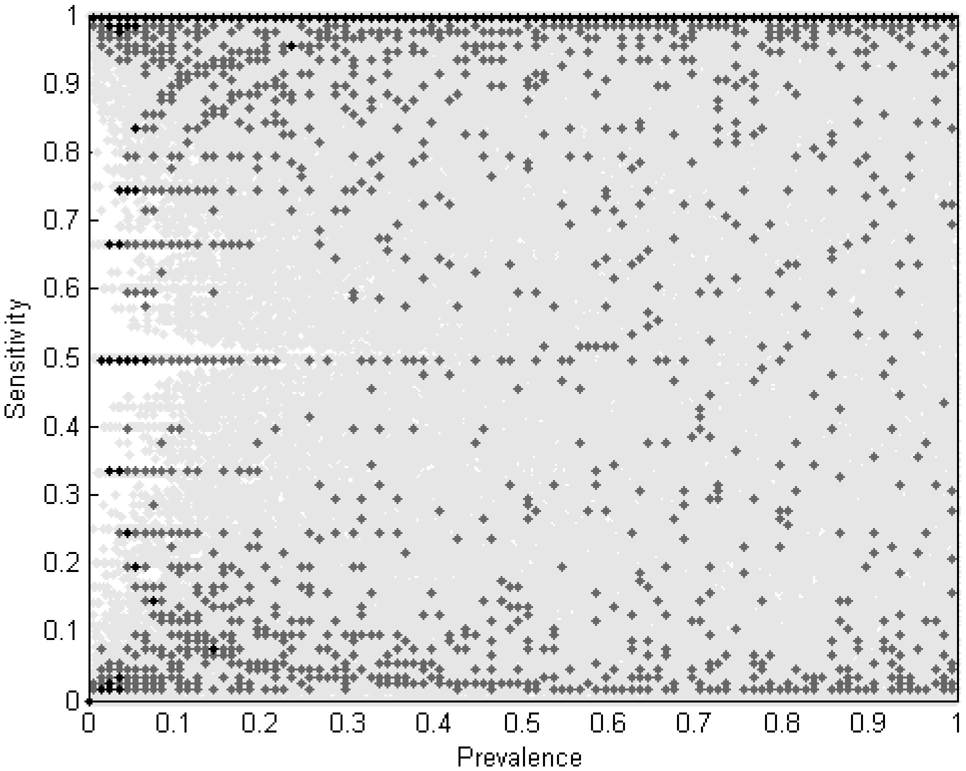


Fig. 6.4. Relation between *Sn* and prevalence, obtained after simulation of 100000 confusion matrices with random prevalence. To reveal patterns in this plot, areas in which the density of confusion matrices exceeded a fixed threshold *x* were selected. The background of this figure is white. Light grey dots indicate all confusion matrices, whereas dark gray and black dots indicate the centre of a 0.01 x 0.01 square in which at least 1 % and 2.5 % of the confusion matrices is situated, respectively.

Plotting the relation between *NMI* and prevalence (Fig. 6.5) reveals that *NMI* shows a weak dependency on prevalence. However, at intermediate prevalence values, perfect *NMI* values only rarely occur. At these prevalence values, fewer matrices occur which contain zeros, and only such confusion matrices can reach maximal *NMI* values due to the assumption that ln(0) equals zero, which is based on the fact that $\lim_{x \to 0} x \ln(x) = 0$. This assumption also explains the gap between the maximal *NMI* values and the other *NMI* values. This weak theoretical dependency of *NMI* to prevalence agrees with empirical analysis of the relation between *NMI* and prevalence (Manel et al., 2001). A similarly weak prevalence dependency is observed for the odds ratio (Fig. 6.7), although lower values occur at extreme prevalence values.

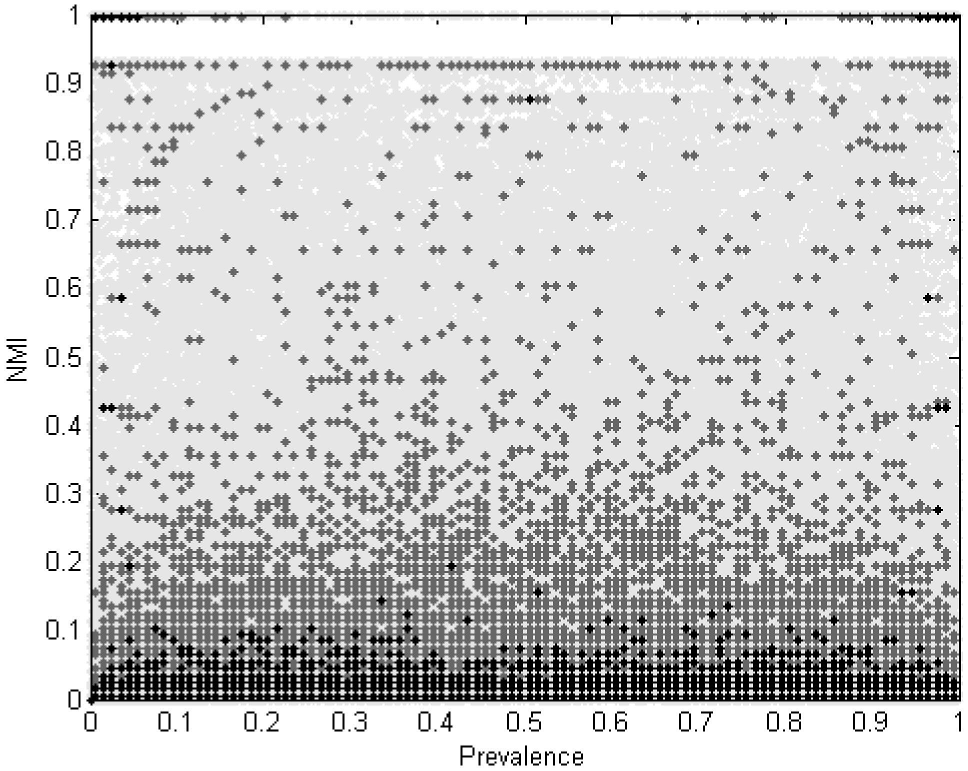


Fig. 6.5. Relation between *NMI* and prevalence, obtained after simulation of 100000 confusion matrices with random prevalence. To reveal patterns in this plot, areas in which the density of confusion matrices exceeded a fixed threshold *x* were selected. The background of this figure is white. Light grey dots indicate all confusion matrices, whereas dark gray and black dots indicate the centre of a 0.01 x 0.01 square in which at least 1 % and 2.5 % of the confusion matrices is situated, respectively.

The positive *Kappa* values in Fig. 6.6 show a similar dependency on the prevalence as *CCI*, with *Kappa* values approaching either 0 or 1 at extreme prevalences. However, the prevalence range in which the density of confusion matrices is similar is wider, whereas for *CCI* this range is centred around the 0.5 prevalence value. Allouche et al. (2006) selected three specific cases from this graph to show the dependency of *Kappa* on prevalence. Yet, this graph provides a wider and more probabilistic view. Each point could be interpreted as the likelihood that, at a certain prevalence level, a specific value of the performance criterion is obtained.

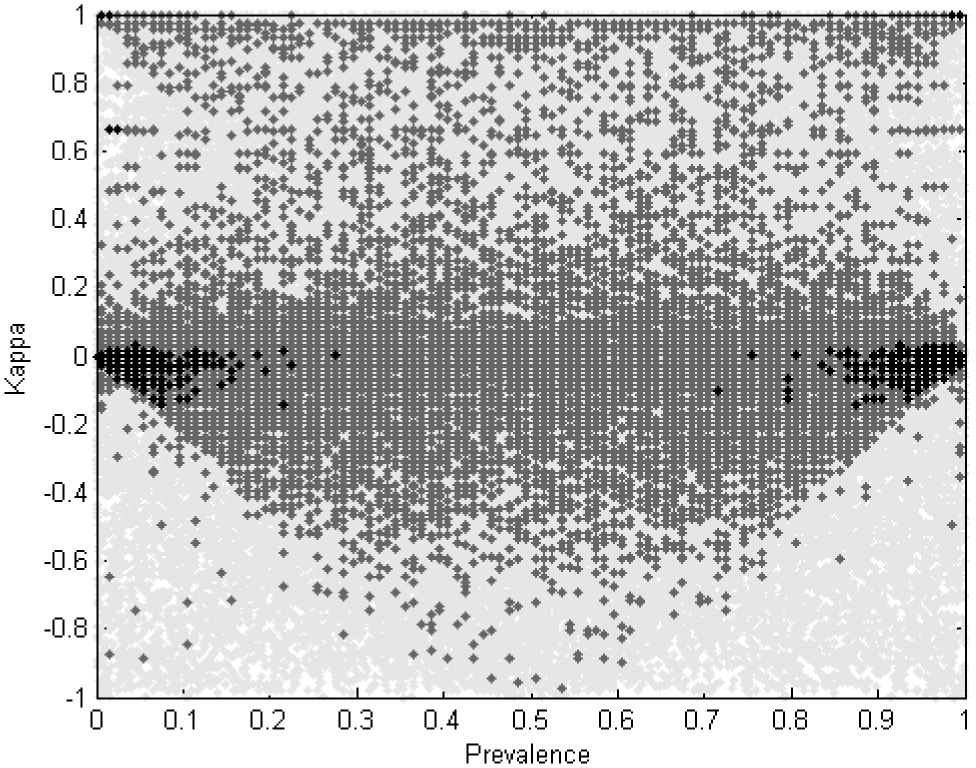


Fig. 6.6. Relation between *Kappa* and prevalence, obtained after simulation of 100000 confusion matrices with random prevalence. To reveal patterns in this plot, areas in which the density of confusion matrices exceeded a fixed threshold *x* were selected. The background of this figure is white. Light grey dots indicate all confusion matrices, whereas dark gray and black dots indicate the centre of a 0.01 x 0.01 square in which at least 0.5 % and 2.5 % of the confusion matrices is situated, respectively.

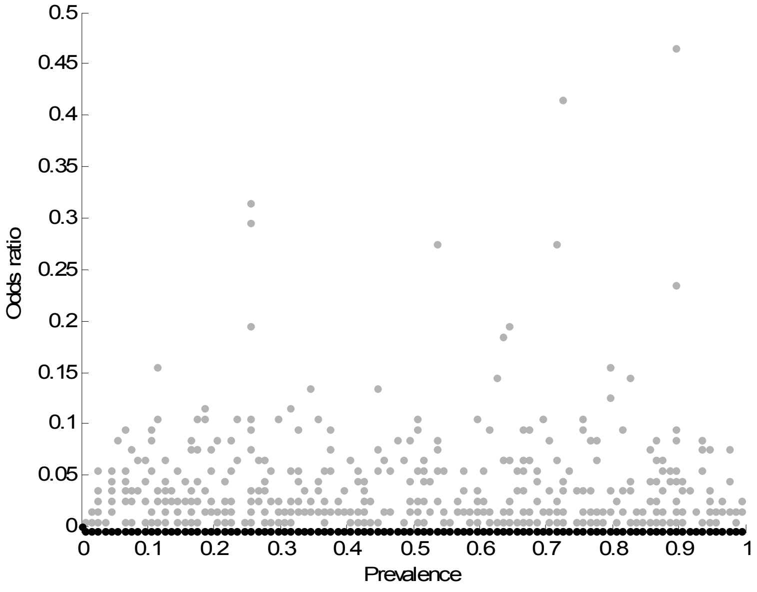


Fig. 6.7. Relation between the odds ratio and prevalence, obtained after simulation of 100000 confusion matrices with random prevalence. To reveal patterns in this plot, areas in which the density of confusion matrices exceeded a fixed threshold *x* were selected. Gray and black dots indicate the centre of a 0.01 x 0.01 square in which at least 0.5 % and 2.5 % of the confusion matrices is situated, respectively. For practical reasons, the odds ratio of all confusion matrices was not displayed because it ranged from 0 to 2500.

Another statistic which was claimed to be independent of prevalence, is the true skill statistic (Allouche et al., 2006). However, plotting the relation between *TSS* and prevalence shows that at extreme prevalence values, *TSS* values either tend to be perfect or extremely low (Fig. 6.8). This theoretical analysis contrasts substantially with the theoretical analysis by Allouche et al. (2006). This might be due to the fact that the latter authors only showed the relation between *TSS* and prevalence for three specific scenarios. However, caution should be made for misinterpretation of both results. The graphs of Allouche et al. (2006) display the likelihood that, at a fixed *Sn* and *Sp* value and at a specific prevalence value, a value of *TSS* is obtained. Fig. 6.8 shows the likelihood that, at a specific prevalence value, independent of the values of *Sn* and *Sp*, a specific *TSS* value is obtained. Moreover, these findings do agree with the results of Allouche et al. (2006) obtained from variance analysis of *TSS*, that showed that *TSS* is highly variable for extremely low and high levels of prevalence. Nelson and Cicchetti (1995) argue that instability at extreme levels of prevalence seems to be inherent in any model with a low number of instances in one of the cells of the confusion matrix.

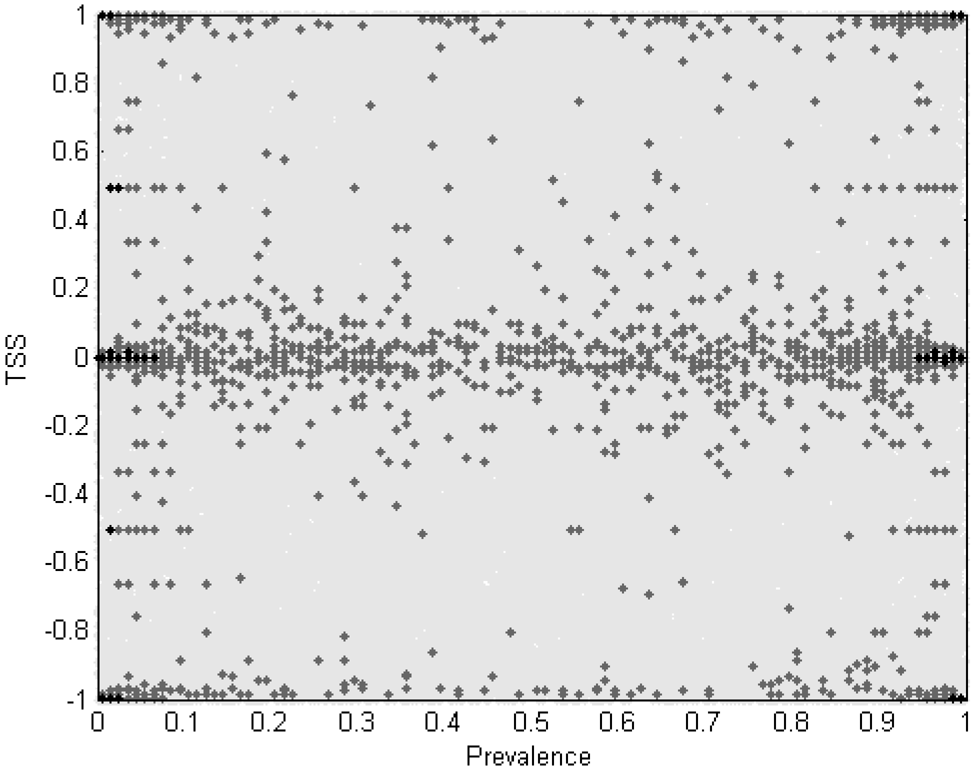


Fig. 6.8. Relation between *TSS* and prevalence, obtained after simulation of 100000 confusion matrices with random prevalence. To reveal patterns in this plot, areas in which the density of confusion matrices exceeded a fixed threshold *x* were selected. The background of this figure is white. Light grey dots indicate all confusion matrices, whereas dark gray and black dots indicate the centre of a 0.01 x 0.01 square in which at least 1 % and 2.5 % of the confusion matrices is situated, respectively.

In the previous section, the relation between the evaluation set prevalence and the performance criteria has been assessed for prevalence values which were uniformly distributed between 0 and 1. In empirical studies, however, not the prevalence but the confusion matrix is randomly selected, which may lead to non-uniform prevalence distributions (Fig. 6.2 and Fig. 6.9). The corresponding prevalence distribution results in prevalence versus performance criterion plots which are much less random than the aforementioned ones. This situation was simulated by generating 100000 random confusion matrices and calculating both prevalence and the performance criteria values from each matrix. The resulting plots are shown in Figs. 6.10-6.15 and show a more distinguished relation between prevalence and the performance criteria than the aforementioned plots. These relations agree more with the empirical results which were reported in various studies (Manel et al., 2001; McPherson et al., 2004; Luoto et al., 2005; Allouche et al., 2006; Bulluck et al., 2006).

Algorithm 6.1. Random generation a confusion matrix

```
$n = 4$
initialise a confusion matrix $C$
initialise a random permutation $P$ of 4
$N = 100$
while $n > 1$
    $n = n - 1$
    select the $P_n^{\text{th}}$ position in the confusion matrix $C$
    if $n = 1$
        assign $N$ to this position
    else
        assign a random number $r$ between 0 and $N$ to this position
        $N = N - r$
    end
    remove $P_n$ from the random permutation $P$
end
```

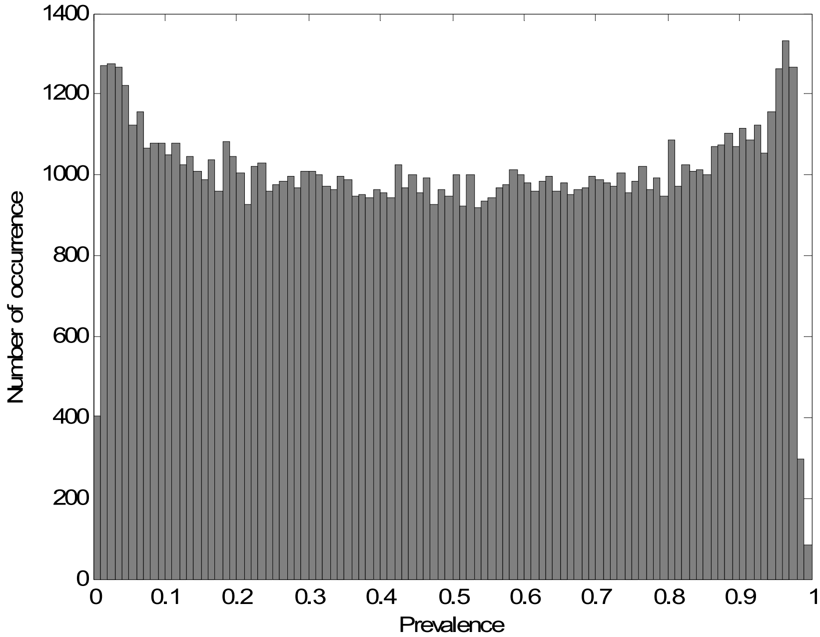


Fig. 6.9. Distribution of prevalence values over 100 prevalence classes after simulation of 100000 random confusion matrices.

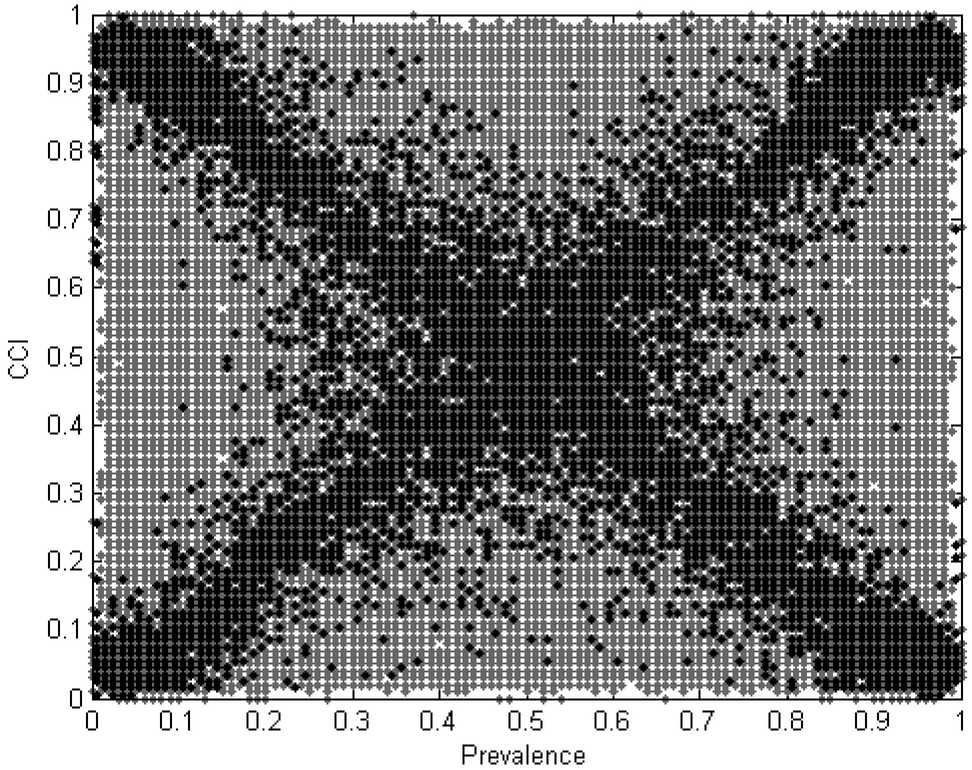


Fig. 6.10. Relation between *CCI* and prevalence, obtained after simulation of 100000 random confusion matrices. To reveal patterns in this plot, areas in which the density of confusion matrices exceeded a fixed threshold *x* were selected. Grey dots indicate all confusion matrices, whereas black dots indicate the centre of a 0.01 x 0.01 square in which at least 1 % of the confusion matrices occurs.

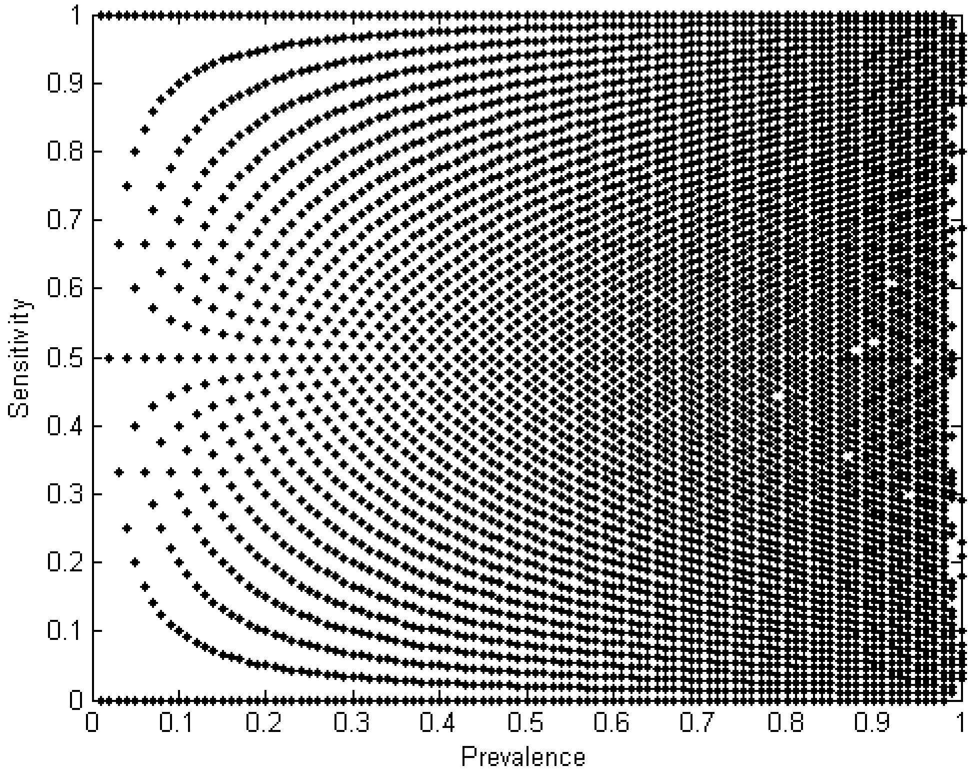


Fig. 6.11. Relation between *Sn* and prevalence, obtained after simulation of 100000 random confusion matrices.

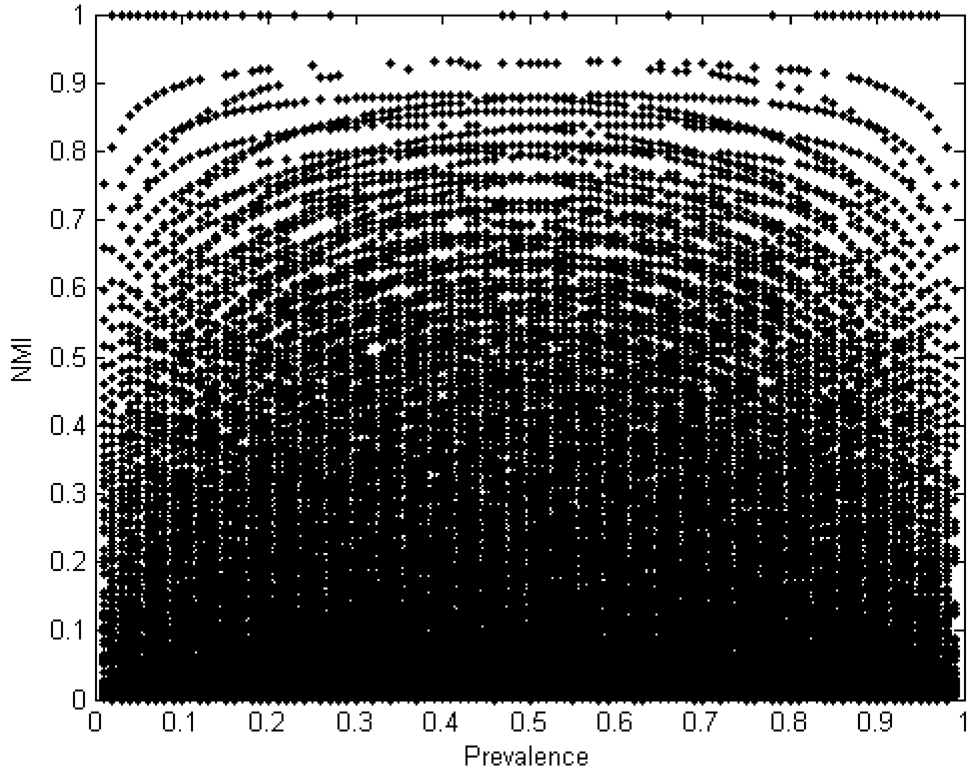


Fig. 6.12. Relation between *NMI* and prevalence, obtained after simulation of 100000 random confusion matrices.

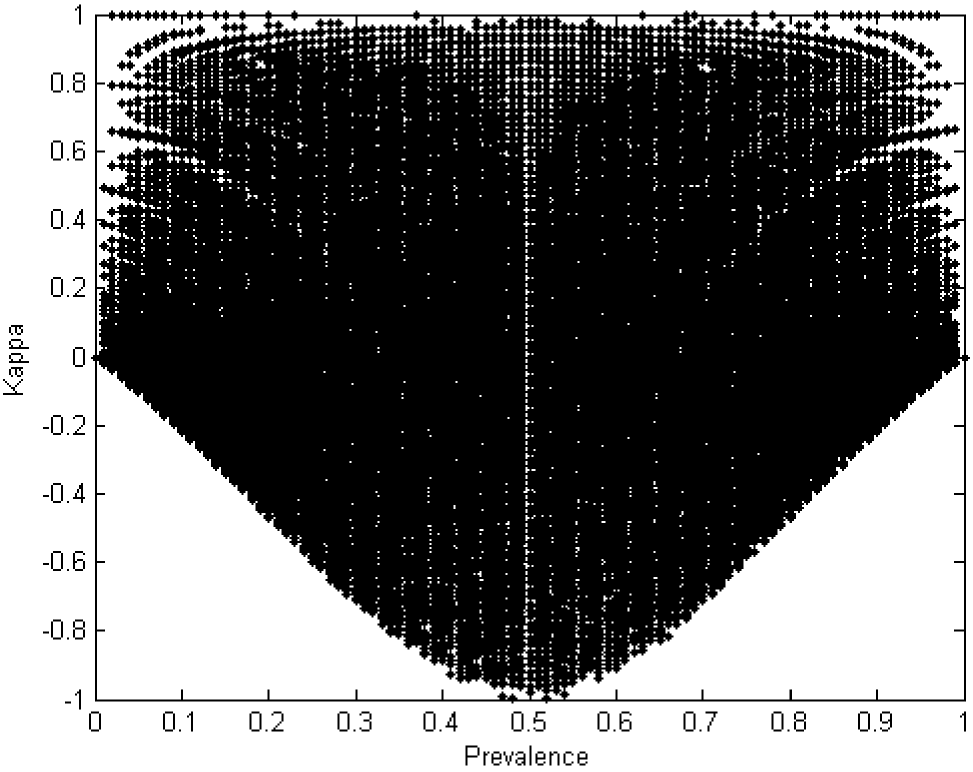


Fig. 6.13. Relation between *Kappa* and prevalence, obtained after simulation of 100000 random confusion matrices.

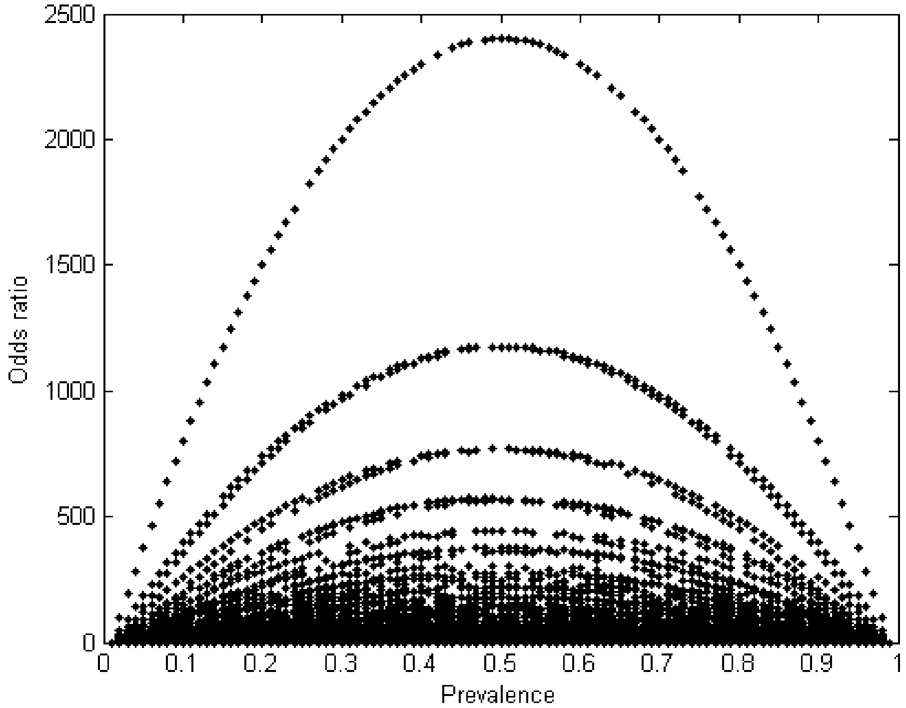


Fig. 6.14. Relation between the odds ratio and prevalence, obtained after simulation of 100000 random confusion matrices.

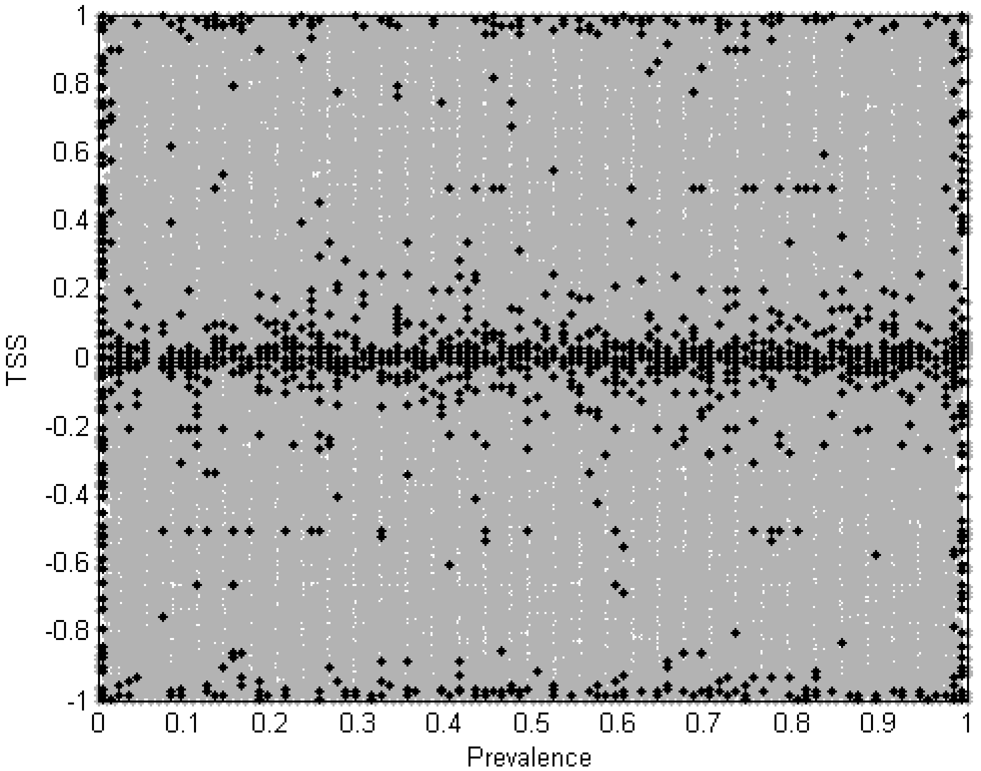


Fig. 6.15. Relation between *TSS* and prevalence, obtained after simulation of 100000 random confusion matrices. To reveal patterns in this plot, areas in which the density of confusion matrices exceeded a fixed threshold *x* were selected. The background of this figure is white. Grey dots indicate all confusion matrices, whereas black dots indicate the centre of a 0.01 x 0.01 square in which at least 1 % of the confusion matrices occurs.

Another important property of performance criteria for model evaluation is the ability to distinguish between omission and commission errors (Loiselle et al., 2003; Wilson et al., 2005). To assess this ability, 4 different scenarios were defined, of which the confusion matrix is given in Table 6.3. In the first scenario, all instances are classified correctly, whereas in the second scenario, all instances are classified erroneously. The third scenario is characterized by slight overprediction, whereas in the fourth scenario, the model is underpredicting the observed data.

Table 6.3. The different scenarios and their corresponding elements of the confusion matrices.

Element of the confusion matrix Scenario	a	b	c	d
1	2	0	0	2
2	0	2	2	0
3	2	1	0	1
4	1	0	1	2

In line with the aforementioned ecological assumptions, an ecologically relevant performance measure would classify scenario 1 as the best scenario and scenario 2 as the worst, while scenario 3 may be considered ecologically more sound than scenario 4 due to the false-negative predictions in the latter scenario. Table 6.4 shows the values of 7 different performance criteria for the 4 scenarios: the percentage of correctly classified instances (*CCI*; Fielding and Bell, 1997), Cohen's *Kappa* (Cohen, 1960), the true skill statistic (*TSS*; McBride and Ebert, 2000; Saseendran et al., 2002; Elmore et al., 2003; Allouche et al., 2006), the *NMI*, the specificity (*Sp*) and the sensitivity (*Sn*). All measures are optimal at their maximum, one, except the odds ratio, which can reach values up to positive infinity.

Table 6.4. The values of the 7 different performance criteria for the 4 scenarios, assuming that ln(0)=0. To calculate the odds ratio, a continuity correction was performed by adding 0.5 to each of the cells in the confusion matrix (Forbes, 1995; Vaughan and Ormerod, 2005). These assumptions have no effect on the characteristics of the presented criteria.

Criterion Scenario	*CCI*	*Sn*	*Sp*	*NMI*	*Kappa*	*TSS*	Odds ratio
1	1	1	1	0	1	1	25
2	0	0	0	0	-1	-1	0
3	0.75	1	0.5	0.23	0.5	0.5	5
4	0.75	0.5	1	0.23	0.5	0.5	5

The performance criteria *CCI*, *NMI*, *Kappa*, *TSS*, and the odds ratio do not distinguish between scenarios 3 and 4, whereas *Sn* and *Sp* allow differentiation between these two scenarios. However, *Sn* and *Sp* do not distinguish between scenario 1, and scenarios 3 and 4, respectively. Because the presented example is based on dichotomous presence-absence model scores, no threshold was applied to generate the confusion matrices. The example illustrates that *TSS* is a special case of *Kappa*, given that the proportions of presences and absences in the training set are equal, which corresponds to a prevalence of 0.5.

6.3.4 Application of performance criteria for model training

The impact of the dependency between the performance criterion and the prevalence differs between model training and model evaluation. During the first process, the prevalence is constant because only one training dataset is used, whereas during model evaluation, prevalence may vary due to the use of different evaluation datasets. Despite the constant prevalence during model training, effects of species prevalence on model training results have been reported (Hirzel et al., 2001). This chapter suggests that the performance criterion used for model training is causing this relation between species prevalence and model training results.

Although numerous papers on performance criteria assessment focus on the application of these criteria in model evaluation (Fielding and Bell, 1997; Manel et al., 2001; Vaughan and Ormerod, 2005; Allouche et al., 2006) or on the choice of an appropriate method for species distribution modelling (Segurado and Araújo, 2004; Elith et al., 2006; McNyset and Blackburn, 2006; Stockman et al., 2006), few authors describe the impact of performance criteria on model training. From the aforementioned group of papers which used data in the modelling process, 99 % applied a training performance criterion which assesses the predictive accuracy of the model, such as *CCI*, Pearson's correlation coefficient, or the root mean squared error. The application of these criteria for model training is based on the assumption that, provided the 'true' model is nested within the model specification, the model estimated using predictive accuracy techniques will converge to the true model as the sample size increases (Welsh, 1996). However, the true model is rarely nested within the model specification due to various factors (Tyre et al., 2001; Barry and Elith, 2006). Moreover, it could be shown theoretically that none of the criteria assessing predictive accuracy distinguish between omission and commission errors. Less than 1% of the evaluated papers applied different performance criteria for model training and compared the results.

The purpose of this section is to analyse the impact of the performance criteria used for model training on the final model which is obtained after training. This impact is based on the two aforementioned characteristics of performance criteria: the dependency on prevalence and the distinction between omission and commission errors. These characteristics are assumed to be independent of each other in model evaluation, whereas this chapter will show that both characteristics are related in model training. Indeed, performance criteria which cannot distinguish between omission and commission error will generate models which underestimate the observations if the prevalence in the training dataset is low, and overestimating models at high prevalence. This section attempts to analyse this relation between prevalence dependency and omission-commission distinction of performance criteria.

Although all criteria allow evaluation of model performance once the model has been developed, some of these criteria are not suitable for model training. An example is the Normalized Mutual Information Statistic (*NMI*; (Forbes, 1995; Manel et al., 2001; Vaughan and Ormerod, 2005), which does not distinguish between the aforementioned scenario 1 and scenario 2 (Table 6.5). Other criteria, such as the *AUC*, are too complex to be efficiently applied for model training. Moreover, *AUC* does not distinguish between omission and commission errors, which can be easily shown in the following example (Table 6.5). Both model B and C are respectively overpredicting and underpredicting one observed value and consequently show the same *AUC* value (Fig. 6.16) of 0.83. Model A shows perfect prediction, which is reflected in an optimal *AUC* value of 1.

Table 6.5. Observed classes and predicted values of three different species distribution models (A = absent; P = present). Model A shows perfect predictions, whereas model B and D are overestimating and model C is underestimating the observations.

Observations	Predictions			
	Model A	Model B	Model C	Model D
A	0	0	0	0
A	0	0	0	0.5
A	0	1	0	1
P	1	1	0	1
P	1	1	1	1
P	1	1	1	1

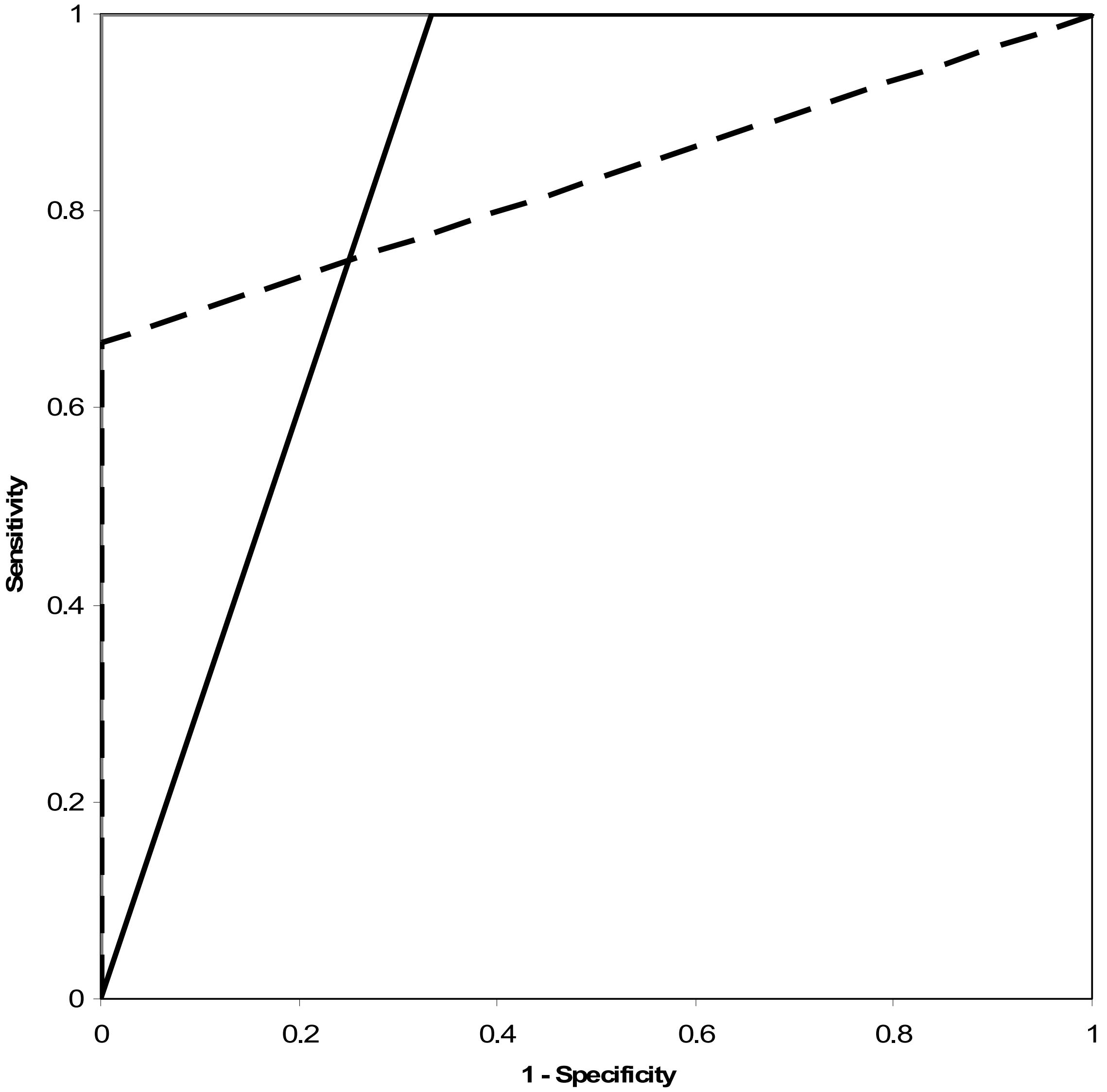


Fig. 6.16. Receiver-operator characteristic curves of the four different species distribution models A (grey line), B (solid black line), C (dashed black line) and D (solid black line).

Although the *AUC* does not distinguish between omission and commission errors, the ROC plot provides some general insight into the tendency of a model to over- or underestimate the observations. Specifically, the ratio of the area under the curve which is above the (0,1) – (1,0) line to the area under the curve below this line gives an indication whether the model is overpredicting or not. A model is generally overestimating the observations if this ratio is greater than 1, and underestimating if this ratio is between 0 and 1. However, this ratio does not indicate the extent to which a model is over- or underpredicting because the shape of the ROC plot is affected by the 'worst' over- or underprediction. This is illustrated by the identical ROC plots of models B and D, although model D is overestimating the observations more than model B. The *AUC* actually indicates the percentage of predicted values that is predicting the observations better that the 'worst' predicted value does.

In this section, three of the most commonly used performance criteria are analysed theoretically: *CCI*, *Kappa* and *TSS*. Specifically, the effect of an increase of the true-positive predictions (a) in the confusion matrix (Δa) on the performance criterion value is assessed. If the training data can be separated into present cases and absent cases by one *N*-dimensional plane, with *N* the number of input variables, the data are referred to as separable. In this chapter, it is assumed that the data are not separable. Consequently, the change of the model parameters which results in a increase of the true-positive predictions Δa, leads to a certain decrease of the true-negative predictions, $-\Delta d$. If the data would be separable, Δd would equal zero. Since prevalence remains constant during model training ($a + c = a^* + c^*$), the confusion matrix will change as described in Table 6.6.

Table 6.6. The confusion matrix after a change of the model parameters which results in an increase Δa of the true-positive predictions a. The table cross-tabulates observed values against predicted values: true-positives, a; false-positives, b; false-negatives, c; true-negative values, d.

		Observed	
		Present	Absent
Predicted	Present	$a + \Delta a$	$b + \Delta d$
	Absent	$c - \Delta a$	$d - \Delta d$

The performance criterion based on the confusion matrix of Table 6.1 will be distinguished from the original performance matrix by an asterisk. Furthermore, the impact of prevalence on the effect of an adjustment of the model parameters which results in Δa will be discussed. The prevalence P of a dataset is defined as the proportion of sites in which the species was recorded as present (Manel et al., 1999a; Allouche et al., 2006) and is described as:

$$P = \frac{a + c}{N} \tag{6.1}$$

Analysis of *CCI*

The change in *CCI*, ΔCCI, which results from a change Δa as described in the confusion matrix in Table 6.6 is equal to:

$$\begin{aligned} \Delta CCI &= CCI^* - CCI \\ &= \frac{(a + \Delta a + d - \Delta d)}{N} - \frac{(a + d)}{N} \\ &= \frac{(\Delta a - \Delta d)}{N} \end{aligned} \tag{6.2}$$

If ΔCCI exceeds zero, the adjustment of the model parameters will result in a better model and the optimisation algorithm will continue with this adjusted model. At low prevalences, Δa tends to be a lot smaller than Δd, whereas at high prevalences, almost all changes of a model parameter will result in a Δa which is greater than Δd. Consequently, the likelihood that ΔCCI will be greater than zero if the prevalence is high is higher than the likelihood if the prevalence is low. This will result in high positive predictions of the model at high prevalences and high negative predictions at low prevalences. At high prevalences, optimisation of *CCI* will thus lead to high overprediction errors, whereas at low prevalences, optimisation of *CCI* will result in high underprediction errors.

Analysis of *Kappa*

To assess the effect of a change Δa as described in the confusion matrix in Table 6.1 on *Kappa*, *Kappa* can be described as a function of the prevalence $P \in \left]0,1\right[$, the sensitivity *Sn* and the specificity *Sp*:

$$Kappa = \frac{P_O - P_e}{1 - P_e} \tag{6.3}$$

with

$$P_O = P \cdot Sn + (1-P) \cdot Sp \tag{6.4}$$

and

$$Pe = -2(Sn + Sp - 1)P(1-P) + P_O\,. \tag{6.5}$$

Consequently, the change in *Kappa*, $\Delta Kappa$, resulting from a change Δa as described in the confusion matrix in Table 6.1 is equal to:

$$\begin{aligned}\Delta Kappa &= Kappa^* - Kappa \\ &= \frac{2(Sn+\Delta Sn+Sp+\Delta Sp-1)\cdot P\cdot(1-P)}{1+2(Sn+\Delta Sn+Sp+\Delta Sp-1)\cdot P\cdot(1-P)-P\cdot(Sn+\Delta Sn)-(1-P)\cdot(Sp+\Delta Sp)} \\ &\quad -\frac{2(Sn+Sp-1)\cdot P\cdot(1-P)}{1+2(Sn+Sp-1)\cdot P\cdot(1-P)-P\cdot Sn-(1-P)\cdot Sp}\end{aligned} \tag{6.6}$$

with

$$\Delta Sn = \frac{\Delta a}{a+c}, \tag{6.7}$$

and

$$\Delta Sp = \frac{\Delta d}{b+d}. \tag{6.8}$$

Eq. (6.6) can be rewritten as:

$$\Delta Kappa = P \cdot (1-P) \cdot \left(\frac{2(TSS) + 2(\Delta Sn + \Delta Sp)}{U} - \frac{2TSS}{V} \right) \tag{6.9}$$

with

$$\begin{aligned} U &= 1 + 2(Sn + \Delta Sn + Sp + \Delta Sp - 1)\cdot P\cdot(1-P) - P\cdot(Sn+\Delta Sn) \\ &\quad - (1-P)\cdot(Sp+\Delta Sp) \\ &= 1 + 2(TSS + \Delta Sn + \Delta Sp)\cdot P\cdot(1-P) - P\cdot(Sn+\Delta Sn) - (1-P)\cdot(Sp+\Delta Sp) \end{aligned} \tag{6.10}$$

and

$$\begin{aligned} V &= 1 + 2(Sn + Sp - 1)\cdot P\cdot(1-P) - P\cdot Sn - (1-P)\cdot Sp \\ &= 1 + 2TSS\cdot P\cdot(1-P) - P\cdot Sn - (1-P)\cdot Sp \end{aligned} \tag{6.11}$$

and thus

$$\begin{aligned} \Delta Kappa &= \frac{P\cdot(1-P)}{U\cdot V}\cdot(2TSS + 4TSS^2\cdot P\cdot(1-P) - 2TSS\cdot P\cdot Sn \\ &\quad - 2TSS\cdot(1-P)Sp + 2\cdot(\Delta Sn+\Delta Sp)\cdot V - 2TSS \\ &\quad - 4TSS^2\cdot P\cdot(1-P) - 4TSS(\Delta Sn+\Delta Sp)\cdot P\cdot(1-P) \\ &\quad + 2TSS\cdot P\cdot(Sn+\Delta Sn) + 2TSS\cdot(1-P)\cdot(Sp+\Delta Sp) \\ &= \frac{P\cdot(1-P)}{U\cdot V}\cdot(2TSS\cdot P\cdot\Delta Sn + 2TSS\cdot(1-P)\cdot\Delta Sp \\ &\quad + 2\cdot(\Delta Sn+\Delta Sp)\cdot V - 4TSS(\Delta Sn+\Delta Sp)\cdot P\cdot(1-P)) \\ &= \frac{2P\cdot(1-P)}{U\cdot V}\cdot(TSS\cdot P\cdot\Delta Sn + TSS\cdot(1-P)\cdot\Delta Sp \\ &\quad + \Delta Sn + \Delta Sp + 2TSS\cdot(P\cdot(1-P)\cdot(\Delta Sn+\Delta Sp)) \\ &\quad - P\cdot Sn\cdot(\Delta Sn+\Delta Sp) - (1-P)\cdot Sp\cdot(\Delta Sn+\Delta Sp) \\ &\quad - 2TSS(\Delta Sn+\Delta Sp)\cdot P\cdot(1-P)) \\ &= \frac{2P\cdot(1-P)}{U\cdot V}\cdot(TSS\cdot P\cdot\Delta Sn + TSS\cdot(1-P)\cdot\Delta Sp \\ &\quad + (1 - P\cdot Sn - (1-P)\cdot Sp)\cdot(\Delta Sn+\Delta Sp)) \end{aligned} \tag{6.12}$$

Since

$$\begin{aligned} (1 - P\cdot Sn - (1-P)\cdot Sp) &= 1 - \frac{a+c}{N}\cdot\frac{a}{a+c} - \frac{b+d}{N}\cdot\frac{d}{b+d} \\ &= \frac{b+c}{N} \quad , \end{aligned} \tag{6.13}$$

and

$$\begin{aligned} TSS\cdot P &= \left(\frac{a}{a+c} + \frac{d}{b+d} - 1\right)\cdot\frac{a+c}{N} = \frac{a}{N} + \frac{d\cdot(a+c)}{N\cdot(b+d)} - \frac{a+c}{N} \\ &= \frac{d\cdot(a+c) - c\cdot(b+d)}{N\cdot(b+d)} = \frac{ad-bc}{N\cdot(b+d)} \quad , \end{aligned} \tag{6.14}$$

and

$$\begin{aligned} TSS\cdot(1-P) &= \left(\frac{a}{a+c}+\frac{d}{b+d}-1\right)\cdot\frac{b+d}{N}=\frac{d}{N}+\frac{a\cdot(b+d)}{N\cdot(a+c)}-\frac{b+d}{N} \\ &= \frac{a\cdot(b+d)-b\cdot(a+c)}{N\cdot(a+c)}=\frac{ad-bc}{N\cdot(a+c)} \quad , \end{aligned} \tag{6.15}$$

Eq. (6.12) can be rewritten as:

$$\begin{aligned} \Delta Kappa &= \frac{2P\cdot(1-P)}{U\cdot V}\cdot(\frac{ad-bc}{N\cdot(b+d)}\cdot\Delta Sn+\frac{ad-bc}{N\cdot(a+c)}\cdot\Delta Sp \\ &\quad +(\frac{b+c}{N})\cdot(\Delta Sn+\Delta Sp)) \\ &= \frac{2P\cdot(1-P)}{U\cdot V}\cdot(\frac{ad-bc}{N\cdot(b+d)}\cdot\frac{\Delta a}{a+c}+\frac{ad-bc}{N\cdot(a+c)}\cdot\frac{\Delta d}{b+d} \\ &\quad +(\frac{b+c}{N})\cdot(\frac{\Delta a}{a+c}+\frac{\Delta d}{b+d})) \\ &= \frac{2P\cdot(1-P)}{U\cdot V\cdot N}\cdot(\frac{\Delta a\cdot(b\cdot(b+d)+d\cdot(a+c))-\Delta d\cdot(a\cdot(b+d)+c\cdot(a+c))}{(a+c)\cdot(b+d)}) \\ &= \frac{2P\cdot(1-P)}{U\cdot V\cdot N}\cdot(\frac{\Delta a\cdot b}{N\cdot P}+\frac{\Delta a\cdot d}{N\cdot(1-P)}-\frac{\Delta d\cdot a}{N\cdot P}-\frac{\Delta d\cdot c}{N\cdot(1-P)}) \\ &= \frac{2P\cdot(1-P)}{U\cdot V\cdot N^2}\cdot(\frac{\Delta a\cdot(b-b\cdot P+d\cdot P)}{P\cdot(1-P)}-\frac{\Delta d\cdot(a-a\cdot P+c\cdot P)}{P\cdot(1-P)}) \\ &= \frac{2}{U\cdot V\cdot N^2}\cdot(\Delta a\cdot((1-P)\cdot b+P\cdot d)-\Delta d\cdot((1-P)\cdot a+P\cdot c)) \end{aligned} \tag{6.16}$$

If $\Delta Kappa$ exceeds zero, the adjustment of the model parameters will result in a better model and the optimisation algorithm will continue with this adjusted model. To assess in which case $\Delta Kappa$ exceeds zero, the sign of U and V is examined for all possible values of Δa and Δd:

$$U = 1+2(TSS+\Delta Sn+\Delta Sp)\cdot P\cdot(1-P)-P\cdot(Sn+\Delta Sn)-(1-P)\cdot(Sp+\Delta Sp) \qquad (6.10$$

and

$$V = 1+2TSS\cdot P\cdot(1-P)-P\cdot Sn-(1-P)\cdot Sp\,. \qquad (6.11$$

Since

$$P\cdot Sn=\frac{a}{N}, \tag{6.17}$$

$$(1-P)\cdot Sp=\frac{d}{N}, \tag{6.18}$$

$$P\cdot\Delta Sn=\frac{\Delta a}{N}, \tag{6.19}$$

$$(1-P)\cdot\Delta Sp=\frac{-\Delta d}{N}, \tag{6.20}$$

Eq. (6.10) and Eq. (6.11) can be rewritten as:

$$U = 1 + 2TSS \cdot P \cdot (1-P) + \frac{2\Delta a \cdot (1-P)}{N} - \frac{2\Delta d \cdot P}{N} - \frac{a}{N} - \frac{\Delta a}{N} - \frac{d}{N} + \frac{\Delta d}{N} \tag{6.21}$$

$$V = 1 + 2TSS \cdot P \cdot (1-P) - \frac{a}{N} - \frac{d}{N} \tag{6.22}$$

Since

$$\begin{aligned} TSS \cdot P \cdot (1-P) &= \left(\frac{a}{a+c} + \frac{d}{b+d} - 1 \right) \cdot \frac{a+c}{N} \cdot \frac{b+d}{N} \\ &= \frac{a \cdot (b+d)}{N^2} + \frac{d \cdot (a+c)}{N^2} - \frac{(a+c) \cdot (b+d)}{N^2} \\ &= \frac{a \cdot (b+d)}{N^2} - \frac{b \cdot (a+c)}{N^2} \\ &= \frac{a \cdot (1-P)}{N} - \frac{b \cdot P}{N} \quad , \end{aligned} \tag{6.23}$$

Eq. (6.21) and Eq. (6.22) can be rewritten as:

$$\begin{aligned} U &= 1 + 2\left(\frac{a \cdot (1-P)}{N} - \frac{b \cdot P}{N} \right) + \frac{\Delta a}{N} - 2\frac{\Delta a \cdot P}{N} - 2\frac{\Delta d \cdot P}{N} - \frac{a+d}{N} + \frac{\Delta d}{N} \\ &= \frac{c+b}{N} + \frac{2a - 2a \cdot P}{N} - \frac{2b \cdot P}{N} + \frac{\Delta a + \Delta d}{N} \cdot (1-2P) \\ &= \frac{a+c}{N} + \frac{(a+b)}{N} \cdot (1-2P) + \frac{\Delta a + \Delta d}{N} \cdot (1-2P) \end{aligned} \tag{6.24}$$

and

$$\begin{aligned} V &= 1 + 2\left(\frac{a \cdot (1-P)}{N} - \frac{b \cdot P}{N} \right) - \frac{a}{N} - \frac{d}{N} \\ &= \frac{a+c}{N} + \frac{(a+b)}{N} \cdot (1-2P) \quad . \end{aligned} \tag{6.25}$$

It can be proven that U always exceeds zero ad absurdum. Indeed,

$$\begin{aligned} U \le 0 \quad &\Leftrightarrow \frac{a+c}{N} + \frac{(a+b)}{N} \cdot (1-2P) + \frac{\Delta a + \Delta d}{N} \cdot (1-2P) \le 0 \\ &\Leftrightarrow P + \frac{a + \Delta a + b + \Delta d}{N} \cdot (1-2P) \le 0 \\ &\Leftrightarrow (N - 2(a + \Delta a + b + \Delta d)) \cdot P + (a + \Delta a + b + \Delta d) \le 0 \end{aligned} \tag{6.26}$$

Now, three cases are distinguished, depending on the sign of the coefficient of P. If the coefficient is strictly positive, i.e.

$$N - 2(a + \Delta a + b + \Delta d) > 0 \Rightarrow \frac{N}{2} > a + \Delta a + b + \Delta d \, , \tag{6.27}$$

then

$$P \le -\frac{a+\Delta a+b+\Delta d}{N-2(a+\Delta a+b+\Delta d)}, \tag{6.28}$$

In this situation, $-\frac{a+\Delta a+b+\Delta d}{N-2(a+\Delta a+b+\Delta d)}$ is always smaller than zero, because a, b, Δa and Δd are positive and not all zero. Consequently, P should be smaller than zero, which is impossible.

If the coefficient of P is strictly negative, i.e.

$$N-2(a+\Delta a+b+\Delta d) < 0 \Rightarrow \frac{N}{2} < a+\Delta a+b+\Delta d, \tag{6.29}$$

then

$$P \ge -\frac{a+\Delta a+b+\Delta d}{N-2(a+\Delta a+b+\Delta d)}, \tag{6.30}$$

In this situation, $-\frac{a+\Delta a+b+\Delta d}{N-2(a+\Delta a+b+\Delta d)}$ is always greater than or equal to one. Indeed,

$$\begin{aligned} -\frac{a+\Delta a+b+\Delta d}{N-2(a+\Delta a+b+\Delta d)} < 1 \quad &\Leftrightarrow a+\Delta a+b+\Delta d < 2(a+\Delta a+b+\Delta d)-N \\ &\Leftrightarrow a+\Delta a+b+\Delta d > N \end{aligned} \tag{6.31}$$

which is impossible since the maximal values of Δa and Δd are c and d, respectively. If $-\frac{a+\Delta a+b+\Delta d}{N-2(a+\Delta a+b+\Delta d)}$ equals one, Δa and Δd should attain their respective maximal values c and d. In this situation, Eq. (6.26) shows that

$$\begin{aligned} U \le 0 \quad &\Leftrightarrow (N-2N)\cdot P+N \le 0 \\ &\Leftrightarrow 1-P \le 0 \end{aligned} \tag{6.32}$$

which is impossible.

The last case to be considered is when the coefficient of P is zero, i.e.

$$N-2(a+\Delta a+b+\Delta d) = 0\,. \tag{6.33}$$

and implies that

$$a+\Delta a+b+\Delta d \le 0, \tag{6.34}$$

which is impossible.

It can be proven that V always exceeds zero ad absurdum. Suppose that

$$\begin{aligned} V \leq 0 \quad & \Leftrightarrow \frac{a+c}{N} + \frac{(a+b)}{N} \cdot (1-2P) \leq 0 \\ & \Leftrightarrow (1-2P) \leq -\frac{a+c}{(a+b)} \\ & \Leftrightarrow 1 - \frac{a+c}{(a+b)} \leq 2\frac{a+c}{N} \\ & \Leftrightarrow 2a \cdot N + b \cdot N + c \cdot N \leq 2a^2 + 2a \cdot c + 2a \cdot b + 2b \cdot c \\ & \Leftrightarrow 2a \cdot (N-a-b-c) + b \cdot (N-c) + c \cdot (N-b) \leq 0 \\ & \Leftrightarrow 2a \cdot d + b \cdot (a+b+d) + c \cdot (a+c+d) \leq 0 \end{aligned} \tag{6.35}$$

The latter inequality is again impossible.

Since U and V are strictly positive, Eq. (6.16) shows that

$$\begin{aligned} \Delta Kappa > 0 \quad & \Leftrightarrow \frac{2}{U \cdot V \cdot N^2} \cdot (\Delta a \cdot ((1-P) \cdot b + P \cdot d) - \Delta d \cdot ((1-P) \cdot a + P \cdot c)) > 0 \\ & \Leftrightarrow \Delta a \cdot ((1-P) \cdot b + P \cdot d) - \Delta d \cdot ((1-P) \cdot a + P \cdot c) > 0 \\ & \Leftrightarrow \Delta a \cdot ((1-P) \cdot b + P \cdot d) > \Delta d \cdot ((1-P) \cdot a + P \cdot c) \\ & \Leftrightarrow \frac{\Delta a}{((1-P) \cdot a + P \cdot c)} > \frac{\Delta d}{((1-P) \cdot b + P \cdot d)} \quad , \end{aligned} \tag{6.36}$$

where the last equivalence follows from the fact that $((1-P) \cdot b + P \cdot d)$ and $((1-P) \cdot a + P \cdot c)$ are strictly positive.

If $\Delta Kappa$ exceeds zero, the adjustment of the model parameters will result in a better model and the optimisation algorithm will continue with this adjusted model. Whether a change in the confusion matrix Δa will lead to a positive $\Delta Kappa$, depends on the prevalence P and on the proportions in the confusion matrix. For instance, if P exceeds 0.5 and $c << d$, a small change Δa will lead to a positive $\Delta Kappa$, even at high Δd values. This relation between $\Delta Kappa$, P and the proportions in the confusion matrix is shown in Table 6.7.

Table 6.7. The effect of the prevalence P and the proportions in the confusion matrix on the likelihood that $\Delta Kappa$ will exceed zero at a fixed increase of the true-positive predictions, Δa. ‘High’ indicates that this relative likelihood is high, while ‘low’ indicates that this likelihood is low.

Prevalence	$a < b$	$a > b$	$c < d$	$c > d$
$P < 0.5$	Low	High	Low	High
$P = 0.5$	See ‘Analysis of *TSS*’			
$P > 0.5$	High	Low	High	Low

Analysis of *TSS*

The change in *TSS*, Δ*TSS*, resulting from a change Δ*a* as described in the confusion matrix in Table 6.1 is equal to:

$$\begin{aligned}\Delta TSS &= TSS^{*} - TSS \\ &= \left(\frac{a+\Delta a}{a+\Delta a+c-\Delta a}+\frac{d-\Delta d}{d-\Delta d+b+\Delta d}-1\right)-\left(\frac{a}{a+c}+\frac{b}{b+d}-1\right) \\ &= \frac{\Delta a}{a+c}-\frac{\Delta d}{b+d}\end{aligned} \tag{6.37}$$

$$\text{Therefore } \Delta TSS > 0 \Leftrightarrow \Delta a > \frac{\Delta d \cdot (a+c)}{b+d} \tag{6.38}$$

If Δ*TSS* exceeds zero, the adjustment of the model parameters will result in a better model and the optimisation algorithm will continue with this adjusted model. If prevalence is greater than 0.5, the difference Δ*a* – Δ*d* should be greater than when prevalence is smaller than 0.5. In the latter case, Δ*TSS* could even exceed zero for some situations in which Δ*a* < Δ*d*. Consequently, a change Δ*a* in the confusion matrix will more easily lead to adjusted model parameters if prevalence < 0.5 and overprediction (*b*) will be stimulated in this situation. At prevalences higher than 0.5, a change Δ*a* in the confusion matrix will only result in a change of the model parameters if Δ*a* substantially exceeds Δ*d*. The optimisation algorithm will thus stimulate underprediction (*c*) (Table 6.1). Comparison of Eq. (6.36) and Eq. (6.38) shows that *TSS* is a special case of *Kappa*, when *P* equals 0.5. Furthermore, the stimulation of either underprediction or overprediction is more complex with *Kappa* than with *TSS*, which is reflected by the more complex denominators in Eq. (6.36) than in Eq. (6.38).

Table 6.8. The relation between Δ*a* and Δ*d* as a function of prevalence. The possible relation between Δ*a* and Δ*d* given is the relation at which Δ*TSS* could exceed 0.

Prevalence	$\frac{a+c}{b+d}$	Possible relations between Δ*a* and Δ*d* such that Δ*TSS* > 0	Stimulation of
$P > 0.5$	> 1	Δ*a* >> Δ*d*	underprediction
$P = 0.5$	=1	Δ*a* > Δ*d*	/
$P < 0.5$	< 1	Δ*a* < Δ*d* or Δ*a* > Δ*d*	overprediction

Table 6.8 shows that the result of model optimisation based on *TSS* depends on the prevalence of the training data set. This emphasises the difference between model training and validation. Although several authors proved that *TSS* is independent of prevalence when it is used for model validation (Allouche et al., 2006), this performance criterion clearly depends on prevalence when it is applied in model training. Specifically, model training based on *TSS* attempts to compensate for the prevalence of the training data: if this prevalence is high, underprediction is stimulated, whereas low prevalences correspond to the stimulation of overprediction.

This issue is of key importance in ecological modelling studies because it shows the effect of the performance criterion used for model training on the resulting model, and thus on the decisions which are supported by this model. At the start of the model development process, modellers should clearly define the goals of the model and then choose a performance criterion which reflects these model purposes.

6.4 Towards prevalence-adjusted model evaluation

The development of a reliable data-driven model requires a sound training and validation procedure. The key component of both procedures is the performance criterion which is applied to evaluate the model performance. Although several authors emphasized the importance of these procedures, model training and validation are often neglected or applied erroneously (Fielding and Bell, 1997; Allouche et al., 2006). This dissertation aims to incorporate both procedures into ecological modelling in an ecologically relevant way.

This chapter showed that different performance criteria evaluate a model (or its resulting confusion matrix) differently. Several authors attributed this difference to the relation between the performance criteria and the prevalence of the validation set (Manel et al., 1999a; Manel et al., 2001; McPherson et al., 2004; Allouche et al., 2006). Specifically, if models derived from different datasets are being compared, the prevalence of these datasets may affect the value of the performance measures and consequently influence the results of the comparison. Similar problems rise when the performances of a model on a training set and a validation set with different prevalences are compared (Allouche et al., 2006). Some authors suggested that this problem would be avoided if validation sets would be collected such that prevalence would be around 50 % (Lantz and Nebenzahl, 1996; Hoehler, 2000; McPherson et al., 2004). However, various authors agree that this recommendation is of questionable practicability in species distribution modelling, particularly for rare species for which a small number of presence data is available (Maclure and Willett, 1987; Mackenzie and Royle, 2005; Allouche et al., 2006). Moreover, an appropriate performance criterion is meant to be a tool for communication (Maclure and Willett, 1987). Consequently, it should be very clear which aspect of the model performance is evaluated, and models should be evaluated consistently, that is, independent of, or taking into account, species prevalence. Given the questionable value of *Kappa* for distribution modelling (Figs. 6.6 and 6.13; McPherson et al., 2004; Allouche et al., 2006), Vaughan and Ormerod (2005) agree that measures other than *Kappa* may thus be preferable to evaluate model predictions.

This chapter also highlighted the relative importance the performance criteria give to omission and commission errors as a possible explanation for the differing evaluation scores among performance criteria for the same model. Theoretical analysis revealed that performance criteria may value a perfect model equally, but yet focus on very different aspects of model performance. An example is the assessment of model discriminatory ability: although all performance criteria will attain their optimum for a model with excellent discrimination, not all criteria adequately quantify the discriminatory ability of a model (Vaughan and Ormerod, 2005). Consequently, model developers should carefully choose an appropriate performance criterion for model evaluation which corresponds to the ecological objectives of the optimised model.

The ecological literature has recognised these problems and the ROC technique in particular has received considerable attention (Elith et al., 2006; Meynard and Quinn, 2007). However, most studies focus on the application of performance measures to model evaluation, whereas relatively few authors describe the role of performance measures in model training. Model training is the process in which the model parameters are iteratively adjusted to increase the agreement between the model predictions and the observations, which are referred to as the training data set. Since this agreement is assessed by the performance criterion, model training aims to optimise the performance criterion value.

The previous section showed that the use of different performance criteria for model training could lead to different optimal models because different performance criteria focus on different aspects of model performance. This may explain differences observed between different modelling techniques reported in previous research. Hirzel et al. (2001) compared generalised linear models (GLM) to ecological niche factor analysis (ENFA) to predict the habitat suitability of an artificial species. They found that both methods produced equivalent results at intermediate prevalence values, whereas the results differed substantially at extreme species prevalence levels (Hirzel et al., 2001). These findings are in line with the substantial variation in *CCI* values at extreme prevalences.

Table 6.9. Characteristics of the most frequently applied performance criteria for model training and evaluation. *NMI* = Normalised Mutual Information Statistic; *TSS* = True Skill Statistic; *AUC* = Area Under the Curve; *CCI* = Correctly Classified Instances; *Sn* = Sensitivity; *Sp* = Specificity; x = the characteristic fully applies to the performance criterion; - = the characteristic does not apply to the performance criterion; ? = the characteristic may apply to the performance criterion.

Performance criterion Characteristic	*Kappa*	*NMI*	Odds ratio	*TSS*	*AUC*	*CCI*	*Sn*/*Sp*
○ quantifies the extent to which models correctly predict occurrence better than chance expectation	x	x	x	x	x	-	-
○ depends on prevalence	x	?	?	x	?	x	x
○ takes into account the complete information included in the confusion matrix	x	x	x	x	n.a.	-	-
○ does distinguish between omission and commission errors	-	-	-	-	-	-	-
○ compensates for extreme prevalence values when applied on model training	x	-	-	-	-	-	-
○ requires discretisation of model predictions by applying threshold values	x	x	x	x	-	x	x
○ allows zero values in the confusion matrix	x	-	-*	x	n.a.	x	x**
○ is proportional (the same performance is found if all elements of the confusion matrix are divided by the same constant)	x	-	-	x	n.a.	x	x
○ is suitable for model training	x	-	x	x	-	x	x

* cannot be applied directly when both the number of false-positive predictions and false-negative predictions is zero; adding a constant value to each element of the confusion matrix changes the relative value of the odds ratio

** cannot be applied if the prevalence of the evaluation or training set is 0.

The effect of the performance criterion on the optimal model depends on the separability of the training data. Training sets which are linearly separable could lead to perfect predictions for all performance criteria, as shown in scenario 1 (Table 6.4). However, in ecological case studies, an increase of the true-positive predictions *a* often results in a decrease of the true-negative predictions *d* and vice versa because the training data are rarely separable. Even with all variables affecting species distribution in the model, prediction is not perfect because of demographic variation (Tyre et al., 2001; Barry and Elith, 2006). The balance between *a* and *d* is affected by the performance criterion which was used to train the model. Consequently, the number of false-positive and false-negative predictions of a data-driven species distribution model depends on this criterion, which was shown theoretically in the previous section. Like for model evaluation, model developers should thus also carefully choose an appropriate training performance criterion which reflects the ecological model purpose

(Segurado and Araújo, 2004). To assist modellers in this choice, Table 6.9 provides an overview of the most important characteristics and restrictions of the most frequently applied performance criteria. The final choice of an appropriate performance criterion is based on a trade-off between different characteristics and depends on the relative importance that modellers attribute to these characteristics.

Different authors suggest to use quantitative predictions whenever possible and to give both threshold independent and dependent measures if nominal predictions need to be derived from quantitative ones (Vaughan and Ormerod, 2005; Barry and Elith, 2006). This approach would provide both a general assessment of performance and one specific to the particular application, threshold and prevalence. However, Vaughan and Ormerod (2005) also point at the dangers of using the quantitative predictions which are produced by many distribution modelling methods. Although such predictions are often considered as estimates of habitat quality or suitability (Buckland and Elston, 1993), this interpretation is only correct if the model predictions show good calibration (Vaughan and Ormerod, 2005). Mackenzie and Royle (2005) also suggest that modellers should take detection probability into account in the statistical inferences. As such, calibration should be considered together with discriminatory ability to allow general interpretation of model results. Instead of assessing calibration in the model evaluation process, good calibration could be obtained by selection of an appropriate performance criterion for model training. Such a criterion should not only focus on the discriminatory ability of the model or on its predictive accuracy, but also on the model calibration. Stimulating calibration during model training avoids the danger of a posteriori misinterpretation of quantitative model results (Vaughan and Ormerod, 2005).

Since calibration is concerned directly with species prevalence (Vaughan and Ormerod, 2005), this performance criterion should also contain a parameter that allows adjustment of the model training process to the training data prevalence. This chapter showed theoretically that training data prevalence may significantly affect the over- or underprediction rate of a model, which is a substantial element of model calibration. Several authors agree that the relative importance of omission and commission errors may vary among applications (Glas et al., 2003; Loiselle et al., 2003; Vaughan and Ormerod, 2005; Wilson et al., 2005). Applying a general performance criterion in model training ignores these subtle but significant differences between different applications. Consequently, an appropriate performance criterion should allow modellers to implement this relative importance in the model training process, for example by including a parameter which can be adjusted to the specific situation.

Although an optimal parameter value could be found by applying sensitivity analysis, a more important problem with these flexible performance criteria could be the difficulty to say which models are better (Glas et al., 2003; Vaughan and Ormerod, 2005). Differences in species' dispersal patterns and associated gene flow may lead to subtle variations in habitat preferences of some species due to local adaptations (Holt, 2003; McPherson and Jetz, 2007). Even in the absence of genetically driven differences in habitat use, species could express different realised niches (Hutchinson, 1957) as a result of spatial variation in predators, competitors or other biotic factors (Hutchinson, 1957; Osborne and Suarez-Seoane, 2002; Holt, 2003; Peterson and Holt, 2003; Hernandez et al., 2006; McPherson and Jetz, 2007). Consequently, conservationists should distinguish between models which reliably or less reliably predict species distribution (McPherson and Jetz, 2007).

The problem of selecting the best model is related to the selection of an optimal threshold value which divides quantitative predictions into two opposing categories. Liu et al. (2005) reported that 4 out of 12 threshold-determining approaches were suitable for reliable threshold selection. All these approaches attempt to minimise both overprediction and underprediction of the model. The prevalence approach and the average suitability approach (Cramer, 2003) adjust the threshold to the training data to avoid over- or underprediction, whereas the sensitivity-specificity equality approach (Cantor et al., 1999) is minimising the distance to the (0,1)-(1,0) line, at which both over- and underprediction of the model are zero. The ROC plot-based approach (Cantor et al., 1999) selects the threshold which corresponds to the point on the ROC curve which has the shortest distance to the top-left corner (0,1) in the ROC plot. This approach corresponds to the previous approach, but also maximises the likelihood that the model predictions differ from chance predictions, which is quantified by the distance to the line y=x.

This may also agree with the results of Maggini et al. (2006), who improved generalised regression analysis for the spatial prediction of forest communities. They found that weighting absences in the training set to bring the average prevalence to a value of 0.5 returned models that seemed to perform better than models fitted with the original sample prevalence. These results show that applying a training parameter or threshold which minimises the difference between over- and underprediction may be a suitable rule of thumb. However, sometimes specific knowledge on model selection is available, for example if omission errors should be considered more unlikely in the application than omission errors. Therefore, it may be more appropriate to apply weights in the training performance criteria than to weight absences in the training data because the first approach allows modellers to adjust the model training process to the purpose to the application. Jiménez-Valverde and Lobo (2006) argue that unbalanced species distribution data are not such a problem from a statistical point of view, and that the effects of unbalanced prevalence should not be confused with those of low-quality data affected by false absences for example. Consequently, it makes more sense to focus on the relation between omission and commission errors during model training than to adjust training data to a prevalence of 0.5. Further research should reveal whether the same model is obtained after threshold optimisation during model evaluation as the model which is obtained by flexible training with a threshold-dependent performance criterion. Another aspect worth investigating is whether weighting absences to ensure a prevalence of 0.5 (Maggini et al., 2006) results in the same model as the model obtained after flexible training. Given the results of Jiménez-Valverde and Lobo (2006), it could be expected that the difference between the results of both approaches is negatively correlated with the quality of the training data.

Given the complexity of the modelled ecological relations, the most robust modelling approaches are likely to be those in which care is taken to match the model with knowledge in ecology. These models should be constrained to be congruent with ecological knowledge, with successive improvement in model performance that is driven by increasing knowledge of the ecology of the system (Leathwick and Whitehead, 2001; Barry and Elith, 2006). Uncertainty in model predictions can thus be viewed from two perspectives: uncertainty as an obstacle that needs to be reduced or removed, or uncertainty as a fact of life (Barry and Elith, 2006). The first approach attempts to change the model structure by seeking more powerful modelling techniques or to improve the data by collecting more samples, removing errors or selecting variables), whereas the second approach tries to understand, characterise and analyse uncertainty by sensitivity analysis, explorations of error or the application of decision strategies that aim to be robust to likely errors (Burgman et al., 2001). Both perspectives are valid and not necessarily mutually exclusive (Barry and Elith, 2006).

6.5 Conclusion

Performance criteria are the key element of the presence/absence model evaluation process and assess the performance of both the final model and the model during training. Although numerous studies on species distribution modelling focus on the role of performance criteria for evaluation of the final model, few authors have addressed the effect of these criteria on model training. This chapter provides a theoretical analysis of the impact of the performance criteria applied for model training on the final model. The results show that, like criteria for evaluation of the final model, the appropriate performance criteria for model training should be chosen carefully and that this choice is dictated by the end-use of the model. For both model training and evaluation, it is suggested that prevalence-independent measures should be preferred, and that at least some of these measures should allow modellers to distinguish between omission and commission errors. The practical implications of this chapter are clear. It provides further insight in the evaluation of ecological presence-absence models and attempts to assist modellers in their choice of suitable performance criteria. As such, it may be an important step towards more reliable species distribution models.

CHAPTER 7

Empirical comparison of performance criteria

7.1 Introduction

In this chapter, a habitat suitability model was developed for spawning grayling in the Aare river (Bern, Switzerland). The model aims at supporting river management in the studied river stretch, which contains some of the major spawning grounds for grayling in Switzerland. To generate an ecologically relevant fuzzy knowledge base for the prediction of spawning grayling presence, a heuristic nearest ascent hill-climbing algorithm was applied. Starting from fixed fuzzy sets, the expert rules were optimised using two different performance measures during training: the weighted % Correctly Classified Instances (% CCI_w) and Cohen's *Kappa* (Cohen, 1960). The aims of this research were: 1) to assess the impact of different performance measures on the optimisation results, 2) to compare the optimised rule base with available expert knowledge and 3) to generate an optimal fuzzy habitat suitability model for spawning grayling.

7.2 Material and Methods

7.2.1 Study area and collected data

The studied site is a 1300 m stretch of the Aare river in the Bern department, Switzerland, and is situated along the city of Thun (Fig. 7.1). Up to this point, the Aare river is draining an area of about 2490 km and is classified as a 7th order stream (Strahler, 1957). The average flow is 111 m /s, with respective base and peak flows of 23 and 570 m /s. The Aare river at the studied site was originally a braided river with large gravel banks. However, since the beginning of the 18th century anthropogenic disturbances were introduced for flood control and hydropower generation (EAWAG, 2002). Hence, the flow regime is altered and controlled by flood control weirs. Nevertheless, the studied site contains some of the major spawning habitats for European grayling in Switzerland.

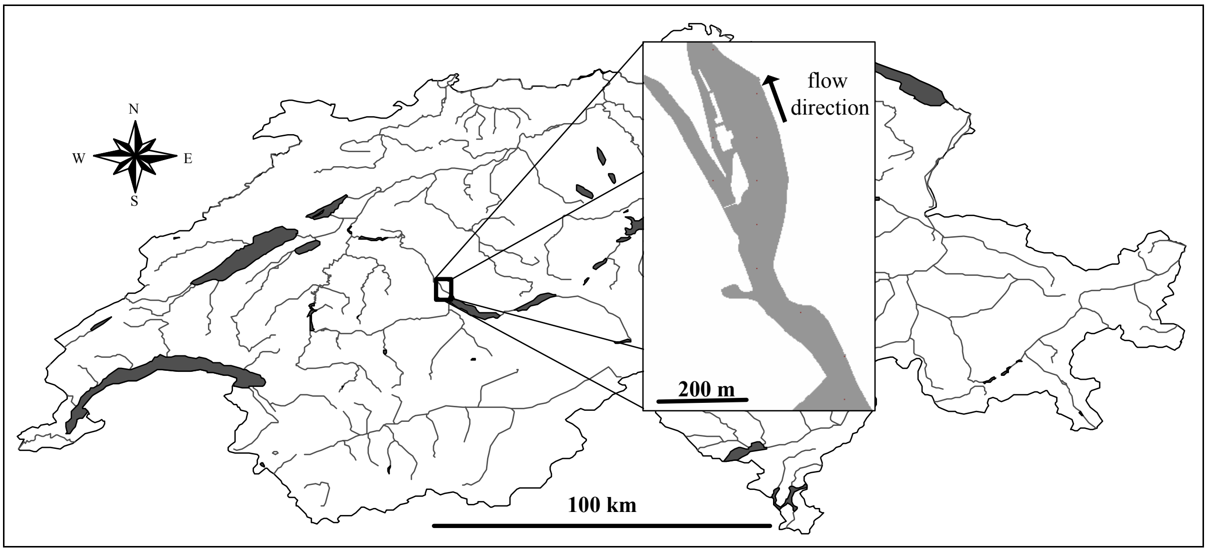


Fig. 7.1. Location of the study site in the Aare river at Thun, Switzerland.

To allow for the development of the species distribution model, an intensive monitoring campaign was set up. In the studied stretch, 50 cross-sections were defined and water depth was measured along each cross-section at equal distances of about 1 m using a Raytheon 760 depth measuring device (Raytheon, MA, USA). Flow velocity was measured with a Flo-Mate 2000 flow meter (Marsh-McBirney Inc., MD, USA) at 40 % of the water column height in 14 of these cross-sections at equal distances of about 25 m, resulting in 63 measurements. The substrate composition of the river bed surface was assessed by underwater photography with DIN A4 frames. Hence, the dominating substrate of the different patches in the studied stretch could be visually assessed. If a patch was covered by macrophytes, both substrate percentages were set to 99 %. This substrate combination can not be observed in the river stretch and hence the definition of this specific situation will not affect optimisation results. All data were collected at a flow of ca. 100 m /s and no significant flow changes were observed during the measurements.

A finite element grid of the studied stretch with 5625 elements and 22500 nodes was generated using SMS (surface water modelling system, Brigham Young University) software, while the size of the grid cells was adjusted depending on river geometry. Flow velocity and depth values were calculated at each node by a 2-dimensional hydraulic model which was generated using FESWMS (Finite Element Surface Water Modelling System, U.S. Geological Survey). Additional measurements of depth and flow velocity were performed in the whole stretch, more specifically in the spawning areas to validate the hydraulic model. The hydraulic modelling was conducted by Schneider & Jorde Ecological Engineering in cooperation with the Swiss Federal Institute of Aquatic Science and Technology (EAWAG) and is described in detail in EAWAG (2002).

European grayling spawns in faster flowing patches (0.1 – 0.4 m/s) with fine to medium-sized gravel substrate. During egg deposition, the trembling female grayling is pushing its abdomen in the gravel substrate, hereby creating small grooves (Fabricus and Gustafson, 1955). These light-coloured grooves can easily be distinguished from the substrate which is mostly covered with dark brown algae. Hence, the spawning grounds of grayling were visually identified and localised using GPS (Garmin 12X). Each grid cell in the studied stretch was defined as suitable for spawning or not by combining the results of the hydraulic simulations with the observations of the spawning grounds. The resulting dataset contained 22510 grid cells, 1 output variable indicating whether the habitat was suitable for spawning or not and 4 input variables characterising each grid cell (Table 7.1).

7.2.2 Fuzzy rule-based modelling and rule base training

Ten-fold cross-validation was applied to indicate the robustness of the optimisation results. The folds were constructed by randomising the original dataset and assigning each data point to one fold without replacement. The species prevalence (i.e. the frequency of occurrence) was constant for all ten folds and equal to the prevalence of the original dataset (0.203 = 4579/22510).

The parameters of the membership functions corresponding to the fuzzy sets of the input variables were derived from an ecological study of spawning grayling in the Aare river (EAWAG, 2002). The parameters of these membership functions are shown in Table 7.1. Since a uniform distribution of the input variables over the fuzzy sets was needed to generate reliable rule bases. The Shannon–Weaver entropy (Shannon and Weaver, 1963) quantified the uniformity and was calculated for the different input variables to assess the quality of the fixed fuzzy sets used in this chapter.

Table 7.1. Input variables recorded and the corresponding fuzzy sets of the species distribution models. The entropy, indicating the uniformity of the distribution of the values of a variable over its fuzzy sets, was calculated for all input variables.

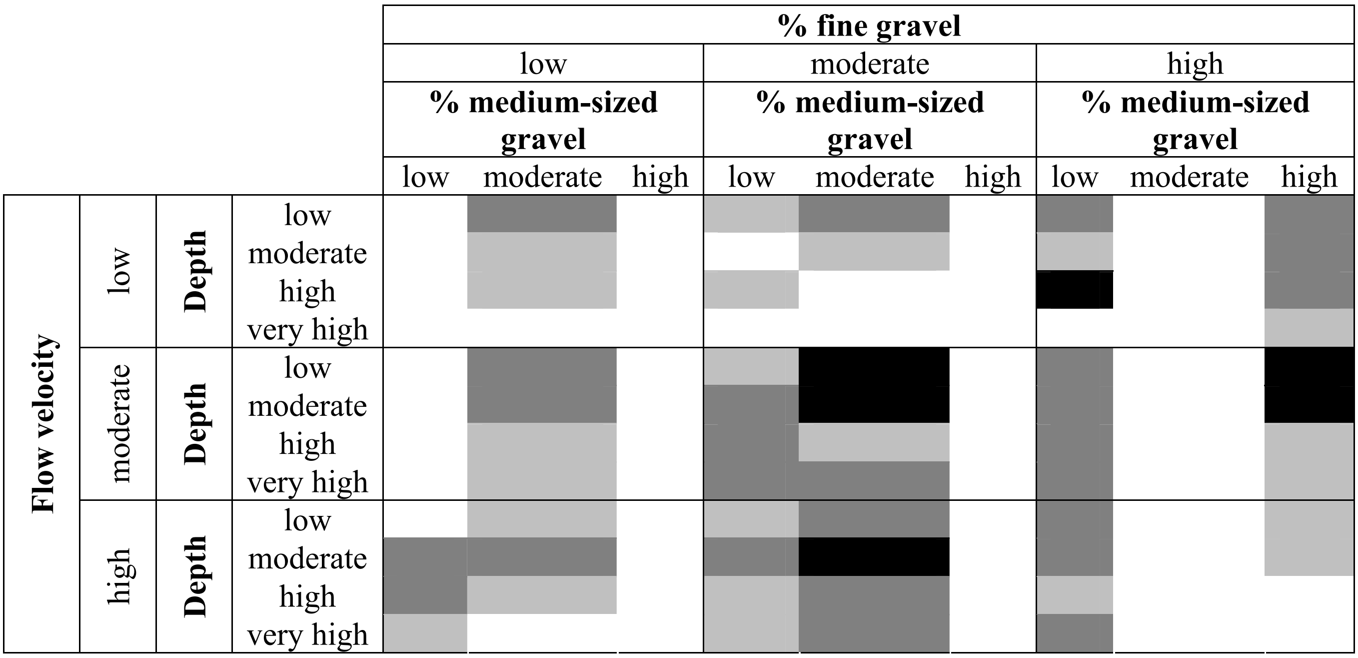
Input variable	**Fuzzy set**	**Parameters**	**Entropy**
Depth (m)	Shallow	(0.0,0.0,0.0, 0.5)	0.808
	Moderate	(0.0,0.5,1.0,2.0)	
	Deep	(1.0,2.0,2.0,3.0)	
	Very deep	(2.0,3.0,6.6,6.6)	
Flow velocity (m/s)	Low	(0.00,0.00,0.05,0.25)	0.818
	Moderate	(0.05,0.25,0.25,0.50)	
	High	(0.25,0.50,0.68,0.68)	
Percentage of fine gravel (2 mm-2 cm) (%)	Low	(0,0,10,50)	0.822
	Moderate	(10,50,50,90)	
	High	(50,90,100,100)	
Percentage of medium-sized gravel (2 cm-5 cm) (%)	Low	(0,0,10,50)	0.940
	Moderate	(10,50,50,90)	
	High	(50,90,100,100)	

Different models were trained based on two performance measures: the percentage of Correctly Classified Instances (Fielding and Bell, 1997), % *CCI*, and Cohen's *Kappa* (Cohen, 1960), both based on the confusion matrix (Fielding and Bell, 1997; Manel et al., 2001) and ranging from zero to one. However, training based on % *CCI* may lead to a high number of false-negative predictions, due to the relatively low species prevalence. To account for this, a weight w was introduced, resulting in the weighted % *CCI* (% CCI_w). The % CCI_w equals the % *CCI* if w, ranging from 0 to 1, is zero. For N data points, the % CCI_w is given by:

$$\% \quad CCI_w = \frac{(a + d + w \times b)}{N} \times 100 \tag{7.1}$$

where a is the number of true-positive predictions, b is the number of false-positive predictions, d is the number of true-negative predictions, w is the weight factor and N is the number of data points. Due to this weight factor, false positive predictions are considered to be true positive predictions. Consequently, overprediction of the observations by the model is stimulated with an increasing value of w. Eleven different training scenarios were created by varying *w* from 0 to 1 with 0.1 steps, while the 12th scenario consisted of *Kappa*-based training (Table 7.2). Each training iteration was stopped when no further increase of the performance measure on the test fold was observed.

Overall model performance (OP) was obtained by multiplying the average score of three performance measures on the test folds: % *CCI*, *Kappa* and *NMI*, the normalised mutual information statistic (Manel et al., 2001). The rule base with the highest overall performance was selected out of the results of the CCI_w optimisation (models 1 to 11). Together with the rule base obtained from *Kappa* optimisation, this rule base was compared with one derived from expert knowledge (EAWAG, 2002). In this chapter, the fuzzy rule base contained 108 (= 3 x 4 x 3 x 3) fuzzy rules representing each possible combination of input variable sets. However, not every combination of input variable sets was present in the studied stretch (Fig. 7.2). To obtain an indication of which rules were relevant, the fuzzy sets were turned into crisp ones by assuming that an input value belongs not to a set if its membership degree to this set is < 0.5. Each data point could be assigned to one environmental condition, resulting in the distribution of the data points over the 108 'crisp' environmental conditions described by the rules (Fig. 7.2). The distribution also gives an indication of the usefulness of the obtained rule base over a range of environmental conditions that can be found in the Aare river.



		% fine gravel								
		low			moderate			high		
		% medium-sized gravel			**% medium-sized gravel**			**% medium-sized gravel**		
Flow velocity	**Depth**	low	moderate	high	low	moderate	high	low	moderate	high
low	low	□	▒	□	░	▒	□	▒	□	▒
	moderate	□	░	□	□	░	□	░	□	▒
	high	□	░	□	░	□	□	■	□	▒
	very high	□	□	□	□	□	□	□	□	░
moderate	low	□	▒	□	░	■	□	▒	□	■
	moderate	□	▒	□	▒	■	□	▒	□	■
	high	□	░	□	▒	░	□	▒	□	░
	very high	□	░	□	▒	▒	□	▒	□	░
high	low	□	░	□	░	▒	□	▒	□	░
	moderate	▒	▒	□	▒	■	□	▒	□	░
	high	▒	░	□	░	▒	□	░	□	□
	very high	░	□	□	░	▒	□	▒	□	□

Fig. 7.2. Distribution of the samples in the dataset over the 108 environmental situations considered in the fuzzy habitat suitability models. All environmental situations were represented by less than 22.9 % of all data points (0 % of instances (□), 0 – 0.5 % of instances (░), 0.5 – 5 % of instances (▒), > 5 % of instances (■)).

7.3 Results

7.3.1 Rule base training

The entropy of all 4 input variables was similar and assumed to be sufficiently high for reliable training of rule bases (Table 7.1). The distribution of instances over the different fuzzy rules was indicated by the entropy of the rule base (0.63). Specifically, this entropy quantifies the uniformity of the distribution of the collected data instances over the environmental conditions included in the rule base. The high entropy values indicated that the fuzzy sets used in this chapter were suitable for rule base optimisation.

Optimisation based on different folds led to the same optimal rule base for all training scenarios and the performance measure values obtained were similar for the training and test folds. Moreover, the *NMI* was higher than the *CCI* and the *CCI* was higher than *Kappa* for each model (Fig. 7.3). Different trends were observed in the values of the three performance measures based on optimisation with changing weights (Fig. 7.3; models 1 to 11).

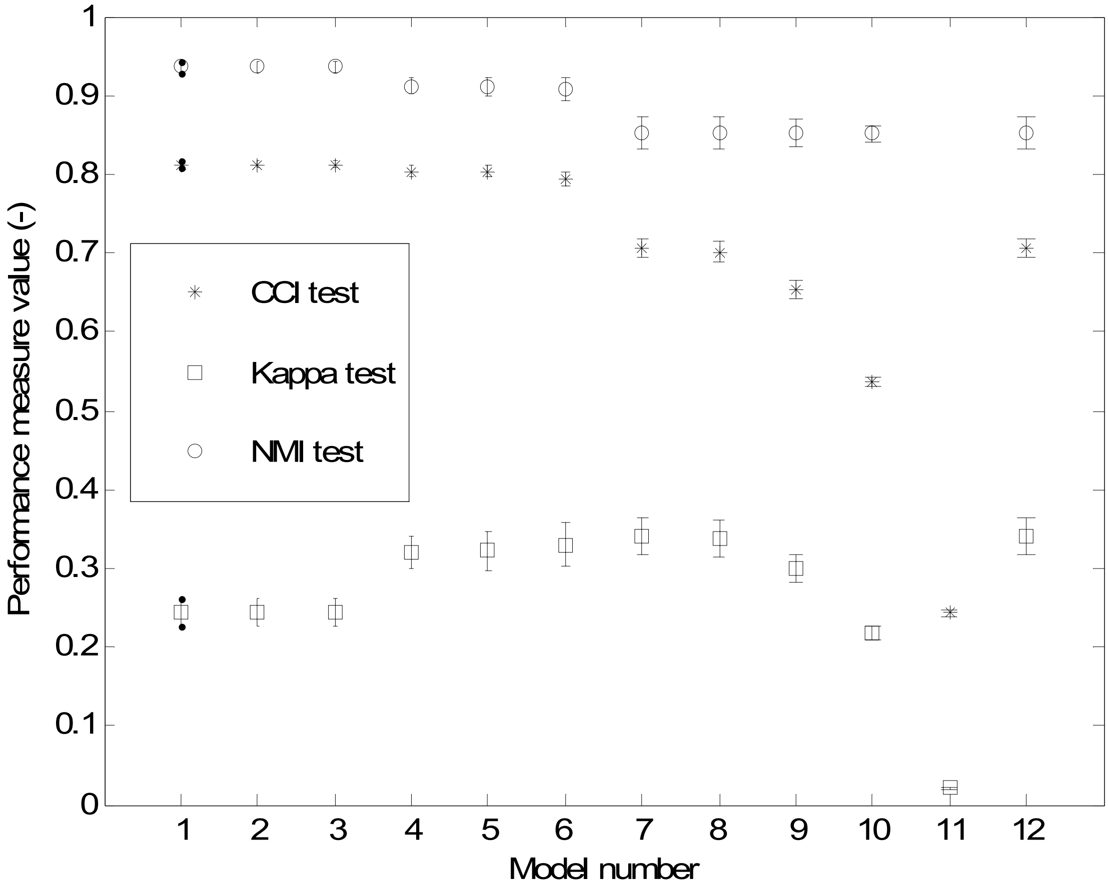


Fig. 7.3. Average performance of the 12 optimised rule bases on classification of the test folds. For all models, the standard deviation of the performance measures based on the test folds is higher then the one of the performance measures based on the training folds, and *NMI* > *CCI* > *Kappa*. *NMI* could not be calculated for rule base 11 as this rule base was predicting all present cases as present, resulting in a c-value of zero in the confusion matrix. (*CCI*= the percentage of correctly classified instances; *NMI*= the normalised mutual information statistic; test= calculated based on the test fold).

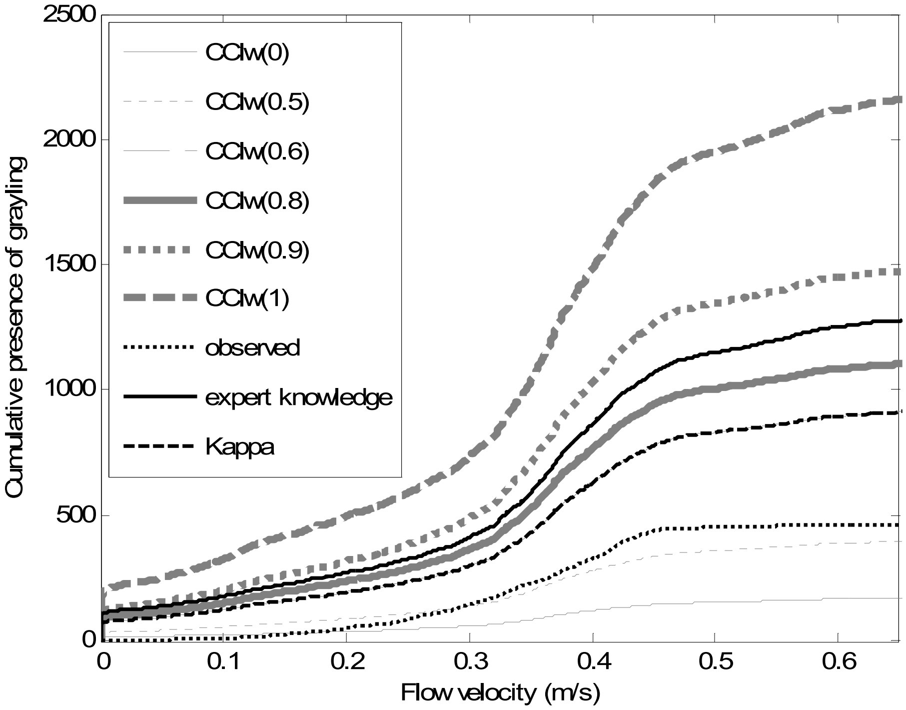
NMI appeared to be relatively unaffected by changing weight, although *NMI* could not be calculated for model 11 ($w = 1$) since this model was predicting all present instances as present, resulting in a zero c-value in the confusion matrix. *Kappa* values were situated along a bell-shaped curve, with its maximum at model 7 ($w = 0.6$). *CCI* was decreasing if w exceeded 0.5 due to the increasing number of false-positive predictions obtained after training at higher weights. For each model, the standard deviation of the performance measures based on the ten folds was lower for the training fold than for the test fold. Standard deviation values were all low (maximum 0.03 and 0.01 on average), indicating that robust rule bases were found after hill-climbing optimisation. Optimisation based on *Kappa* (model 12) led to very similar results as optimisation based on $\mathrm{CCI}_{0.6}$. Furthermore, the values of the three performance measures for this model were particularly high with respect to those resulting from the other optimisation methods.

The overall performance (OP) of the model on the three performance measures was analysed by multiplying the average values of these measures based on the ten test folds (Table 7.2). This indicated that optimisation based on $\mathrm{CCI}_{0.5}$ produced the rule base with the highest OP, while optimisation based on $\mathrm{CCI}_{0.3}$ and $\mathrm{CCI}_{0.4}$ generated similar results. The optimal rule base obtained after running all optimisation scenarios based on $\mathrm{CCI}_{0.5}$ (rule base 6) was selected, together with the rule base obtained after *Kappa* optimisation (rule base 12).

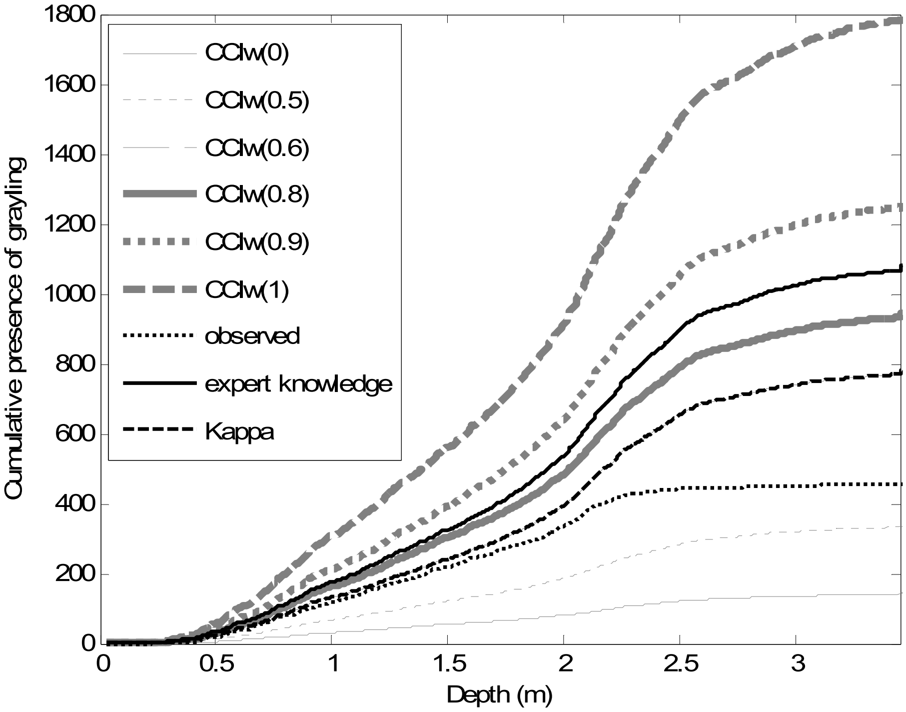
7.3.2 Ecological relevance of the selected rule bases

To assess the ecological relevance of the rule bases obtained after training, the resulting predictions were further investigated. The cumulative predicted presence, based on the 12 different rule bases obtained from each training scenario, was plotted for all variables and displayed for flow velocity and depth (Fig. 7.4). Concerning the location of these cumulative distribution curves, similar results were found for all variables (e.g. Fig. 7.4). When cumulative distribution curve A was situated below curve B for flow velocity for instance, this was also observed for the other three input variables. Results were also similar for all test and training folds and for the whole dataset.

To get insight into the position of the different curves in relation to each other, the example of flow velocity, averaged over the ten test sets, was focussed on (Fig. 7.4a). The cumulative curve of the observed presences was located below the curve of the presences predicted by the expert knowledge rule base. Training based on CCI_{0} resulted in a rule base predicting the most instances as absent. An increasing weight w led to more predictions of present instances, but the prediction evolved discontinuous when w was changing. Predictions of models obtained from training based on $\mathrm{CCI}_{0.1}$, $\mathrm{CCI}_{0.2}$, $\mathrm{CCI}_{0.3}$ and $\mathrm{CCI}_{0.4}$ were identical. If w exceeded 0.8, rule bases obtained from CCI_w training were increasingly predicting present instances (Fig. 7.4a). The CCI_{1} rule base was predicting all present instances as present, leading to low *CCI* and *Kappa* values (Fig. 7.3).



(a)



(b)

Fig. 7.4. Cumulative presence of grayling for flow velocity (a) and depth (b), resulting from observations (observed), expert knowledge and model predictions. Predictions were done by rule bases trained on different performance measures: CCI_0 (= *CCI* or *CCIw*(0)), $CCI_{0.5}$ (= *CCIw*(0.5)), $CCI_{0.6}$ (= *CCIw*(0.6)), $CCI_{0.8}$ (= *CCIw*(0.8)), $CCI_{0.9}$ (= *CCIw*(0.9)), CCI_1 (= *CCIw*(1)) and *Kappa*. To improve visualisation, the depth range between 3.45 and 6.60 m is not shown.

Comparison of performance measures obtained by optimisation based on *Kappa* with the ones obtained by optimisation based on $\mathrm{CCI}_{0.6}$ indicated both optimisation scenarios led to very similar rule bases. The slight differences in the crisp outputs of both rule bases disappeared when these outputs were transformed into present/absent classes for performance measure calculation. The $\mathrm{CCI}_{0.5}$ rule base was approximating the observed curve the closest, but was still showing a substantial number of false-negative predictions. In contrast, the rule bases derived from optimisation based on *Kappa* or $\mathrm{CCI}_{0.6}$ produced a cumulative prediction curve which was situated between the observed curve and the expert knowledge curve.

The ecological significance of the model results could be derived from the gradient of the cumulative presence curves (Fig. 7.4). This gradient indicated where most spawning grayling was observed or predicted, and therefore could be used as measure of the habitat preferences of grayling. For flow velocity (Fig. 7.4a), the gradient of the observed grayling curve was the highest between 0.27 and 0.43 m/s. The highest gradient of the curve derived from expert knowledge was observed in a slightly higher range between 0.31 and 0.45 m/s. This range was similar to the range in which the highest gradients were found of the cumulative curves derived from all model results. For depth (Fig. 7.4b), most spawning grayling was observed between 1.9 and 2.3 m. The depth preferences based on expert knowledge were situated between 1.9 and 2.5 m, while preferences derived from all model results were within the same range. The cumulative presence curves of the two substrate variables showed that spawning grayling preferred coarser substrate. Specifically, the highest gradient of the cumulative presence for the percentage of medium-sized gravel was found between 50 and 70 %, while for the percentage of fine gravel, this ranged between 0 and 40 %. These ranges were very similar for observed data, expert knowledge and model results.

Table 7.2. Overall performance of the 12 rule bases obtained after different training scenarios. During training, a rule base is optimised based on its predictive performance, which is assessed by a different performance measure for each scenario. For each optimal rule base, its overall performance is obtained by multiplying its scores on three performance measures: *CCI*, *NMI* and *Kappa*, all ranging from 0 to 1.

Scenario number	Performance measure applied for training	Weight	Overall Performance
1	$\mathrm{CCI}_0 = CCI$	0.0	0.19
2	$\mathrm{CCI}_{0.1}$	0.1	0.19
3	$\mathrm{CCI}_{0.2}$	0.2	0.19
4	$\mathrm{CCI}_{0.3}$	0.3	0.23
5	$\mathrm{CCI}_{0.4}$	0.4	0.24
6	$\mathrm{CCI}_{0.5}$	0.5	0.24
7	$\mathrm{CCI}_{0.6}$	0.6	0.21
8	$\mathrm{CCI}_{0.7}$	0.7	0.20
9	$\mathrm{CCI}_{0.8}$	0.8	0.17
10	$\mathrm{CCI}_{0.9}$	0.9	0.10
11	CCI_1	1.0	-
12	*Kappa*	-	0.21

7.4 Discussion

In the presented study, the habitat preferences derived from field observations or expert knowledge were very similar to those predicted by the fuzzy rule-based models. Since only few organisms were observed at a depth higher than 2.5 m and a flow velocity higher than 0.45 m/s, model predictions above these thresholds will be less reliable, as was discussed earlier (Fig. 7.2). The predicted flow velocity preferences for spawning grayling were slightly higher than the 0.1 – 0.4 m/s range found by Fabricus and Gustafson (1955). This was confirmed by Nykänen and Huusko (2002), who found flow velocity preferences between 0.5 and 0.6 m/s and suggested that these preferences might be transferable between rivers. The model results were also supported by the wider flow velocity preferences (0.2 – 0.9 m/s) observed in other studies (Gönczi, 1989; Sempeski and Gaudin, 1995).

Several authors describe depth preferences of spawning grayling which are significantly lower (0.1 – 0.5 m) than both the observed and modelled preferences in this study (Gönczi, 1989; Sempeski and Gaudin, 1995; Nykänen and Huusko, 2002). However, these deeper spawning habitats in the Aare river have been observed for more than 20 years (EAWAG, 2002). Possible explanations for the contrasting depth preferences may be differences in topology of the sampled rivers or ecological factors such as competition or predation. Specifically, the shallower spawning areas in the Aare rivers are regularly disturbed by swans, which might cause a shift in depth preferences (EAWAG, 2002).

Unlike the depth preferences, the modelled and observed substrate preferences strongly correspond to previous research, since several authors agree that spawning grayling prefers medium-sized gravel (Gönczi, 1989; Sempeski and Gaudin, 1995; Nykänen and Huusko, 2002). Gönczi (1989) suggests that gravel thickness may also affect spawning preferences. Although this variable was not considered in this chapter because field data were lacking, the presented fuzzy rule-based approach would allow fast and easy implementation of such expert knowledge in the fuzzy knowledge base.

The expert knowledge rule base (EAWAG, 2002) predicted spawning grayling to be present for several instances where no spawning grayling was observed (Fig. 7.4). However, the expert knowledge in this rule base was based on the judgment of local fish experts and is similar to previous findings (Gönczi, 1989; Sempeski and Gaudin, 1995; Nykänen and Huusko, 2002; Nykänen et al., 2004; Riley et al., 2006). Hence, false-positive prediction of the observed situation by the expert knowledge rule base may not necessarily imply a model error. Several factors suggest these false-positive predictions are inherent to the classification of ecological data as discussed in Chapter 6.

Although few hypotheses can explain false-negative predictions by the optimised rule bases, the acceptance of false-positive predictions should be limited, because models predicting spawning grayling to be present at all sites are ecologically irrelevant (Manel et al., 2001). The expert knowledge rule base could be an appropriate limit of acceptable false-positive predictions since it is based on literature. The cumulative predicted presence curve of any correct solution (Fig. 7.4) should thus be situated in the area delimited by the cumulative curves of the expert knowledge model and of the observed presence of grayling. Hence, the four correct models which could be selected in this chapter were the models obtained after optimisation based on *Kappa*, $\mathrm{CCI}_{0.6}$, $\mathrm{CCI}_{0.7}$ and $\mathrm{CCI}_{0.8}$. If the weight w used in CCI_w optimisation was lower than 0.6, the resulting rule bases are all showing a substantial number of false-negative errors due to the dependency of the % CCI_w on the prevalence of the dataset (Manel et al., 2001). For instance, the rule base obtained after training based on $\mathrm{CCI}_{0.5}$ had the highest OP but it was showing a substantial number of false-negative predictions and therefore it may not be an optimal habitat suitability model. Although the weight w compensated the low prevalence of the training folds, values of w exceeding 0.8 led to overcompensation and a questionably high number of false-positive predictions. Further research and validation with new data could point out which model is the most accurate. Model accuracy might also be improved by applying different weights during CCI_w training, but this could lead to similar results due to the discontinuous evolution of model predictions with changing weights.

Comparison of the results from different optimisation scenarios and folds indicated the hill-climbing method generated consistent results. This consistency is enhanced by the size of the dataset used with respect to the relatively small sampling area and by the uniform distribution of the training instances over the different rules. However, Fig. 7.2 indicates that some rules apply to environmental conditions which are not present in the dataset, which was also reflected in the lower entropy of the rule base compared to the entropy of the individual input variables. Since the consequent of these rules cannot be reliably predicted, the optimised rule bases only apply to environmental conditions which are covered by the data (Van Broekhoven et al., 2006).

Optimisation based on *Kappa* resulted in the same cumulative plot as optimisation based on $CCI_{0.6}$. Specifically, the rule bases and the crisp predictions were slightly different but the final fuzzy output was the same. Similarity between both rule bases was indeed increasing from 75 % to 100 % if the number of rules included in the similarity calculation was reduced from 108 to 54. This led to slight differences between the predictions of both models, but these differences disappeared when the predictions were transformed into present/absent classes. Analysis of $CCI_{0.6}$ versus *Kappa* did not reveal any clear relation between both measures or between both performance measures and global data characteristics. Further research should point out if relations between different performance measures exist and indicate how these could be taken into account during the optimisation process.

Overall, the results suggest that the performance measure used for training has a significant impact on the resulting optimised rule base (Fig. 7.3). The results indicate *Kappa* might be a better option for rule base training then CCI_w as the latter requires fine tuning of the w parameter. Hence, further research should reveal which performance measure is the most appropriate for model training, depending on the classification problem and on the model purpose. The presented selection of an appropriate performance measure for model training based on the cumulative predicted presence of this model might also be applied to training of other data-driven models such as artificial neural networks (Lek and Guégan, 1999), decision trees (Džeroski, 2001) and Generalised Linear Models (Brotons et al., 2004).

The inconsistent performance measure values evaluating the optimised models (Fig. 7.3) also indicate that different measures should be compared to evaluate model performance after training, which is reflected in previous research findings (Manel et al., 2001; Hirzel et al., 2006). In line with the performance measure selection for model training, the optimal performance measure for model evaluation should be chosen depending on the classification problem and on the model purpose. The presented case suggests *Kappa* might be an appropriate measure to evaluate model performance, which is supported by the acceptance of *Kappa* as a model performance measure by several authors (Manel et al., 2001; Hirzel et al., 2006). These results also reinforce the view that categorisation of performance indicated by *Kappa* into classes from fair to almost perfect is arbitrary. Hence, the value of *Kappa* should be interpreted by comparison with other *Kappa* values derived from models classifying the same dataset (Manel et al., 2001).

CHAPTER 8

Prevalence-adjusted fuzzy model training

8.1 Introduction

The literature review in Chapter 6 indicated that most studies apply a training performance criterion which assesses the predictive accuracy of the model, such as the percentage of correctly classified instances *CCI* (Fielding and Bell, 1997) or the root mean squared error. However, Chapter 7 showed that model training based on different performance criteria may result in different final models, while the prevalence of the training set (i.e. the frequency of occurrence) may affect the training performance criterion and hence the results of the model training. Specifically, Chapter 6 showed theoretically that model training based on *Kappa* at low prevalences tends to result in models that overestimate the observations, whereas high prevalences may lead to underestimating models.

Consequently, modellers should compare the results of various model training procedures based on different performance criteria and then select the final model which corresponds the most to the study objectives, as was suggested in Chapter 7. Although models could be trained differently by varying the training performance criterion, another option is to train models based on a prevalence-adjusted performance criterion, which could be adjusted to the training data prevalence and to the model objectives. In the previous chapter, a first prevalence-adjusted performance criterion, the CCI_w, was introduced. However, this criterion does not take into account the specific fuzzy characteristics of a fuzzy species distribution model, which will be illustrated in this chapter. Therefore, this chapter presents the adjusted average deviation (*aAD*), which is similar to the average deviation *AD* (Van Broekhoven et al., 2007) but contains a parameter α to account for different prevalences of the model training set. Previous research showed that the *AD* is an appropriate performance measure for fuzzy models because it takes into account the fuzzy characteristics of the model output (Van Broekhoven et al., 2007). In contrast to the *aAD* however, the *AD* does not distinguish omission from commission errors and does not allow prevalence-adjusted model training.

This study aims to analyse the strengths and weaknesses of the *aAD* for ecological modelling. First, the relation between the training set prevalence and the *α* parameter of the *aAD* and the impact of this prevalence on the results of model training based on the *aAD* was analysed by developing models from training sets with different prevalences. The final models obtained after training based on the *aAD* were then compared with those obtained from training based on more frequently applied performance criteria. Finally, the presented procedure was applied on ten artificial datasets to assess the consistency of the dependency between the training set prevalence and the *α* parameter of the *aAD*.

8.2 Material and methods

8.2.1 Study area and collected data

The fuzzy model quantified the habitat suitability for spawning European grayling (*Thymallus thymallus* L.) in a 1300 m stretch of the Aare river in the Bern department, Switzerland. The study area, the sampling procedure and the resulting dataset were described in Chapter 7.

8.2.2 Fuzzy rule-based modelling and rule base training

The parameters of the membership functions corresponding to the fuzzy sets used in this work are given in Table 8.1 and were optimised to create a uniform distribution of the input variables over the fuzzy sets. The Shannon–Weaver entropy (Shannon and Weaver, 1963) quantified this uniformity and was used as an optimisation criterion to increase the quality of the fuzzy sets as described in Chapter 4. The fuzzy rule-based modelling procedure was described in Chapters 4 and 7.

Table 8.1. The input variables and the corresponding fuzzy sets of the species distribution models. The entropy, indicating the uniformity of the distribution of the values of a variable over its fuzzy sets, was calculated for all input variables.

Input variable	**Fuzzy set**	**Parameters**	**Entropy**
Depth (m)	Shallow	(0.0,0.0,0.82, 2.46)	0.94
	Moderate	(0.82,2.46,3.26,3.33)	
	Deep	(3.26,3.33,6.59,7.41)	
Flow velocity (m/s)	Low	(0.00,0.00,0.17,0.50)	1
	High	(0.17,0.50,0.67,0.83)	
Percentage of fine gravel	Low	(0,0,40,60)	0.99
(2 mm-2 cm) (%)	High	(40,60,100,100)	
Percentage of medium-sized gravel	Low	(0,0,30,69)	1
(2 cm-5 cm) (%)	High	(30,69,99,100)	

8.2.3 Training performance criteria

Different models were trained based on five performance criteria: the percentage of Correctly Classified Instances *CCI* (Fielding and Bell, 1997), Cohen's *Kappa* (Cohen, 1960), the true skill statistic *TSS* (Allouche et al., 2006), the average deviation *AD* (Van Broekhoven et al., 2007) and the adjusted average deviation *aAD*. The first three criteria are based on the confusion matrix (Fielding and Bell, 1997; Manel et al., 2001) and range from zero (*CCI* and *Kappa*) or minus one (*TSS*) to one. Although the *TSS* has been applied less frequently than *CCI* and *Kappa*, it was included in this chapter because Allouche et al. (2006) argue that this criterion is independent of prevalence.

The *AD* (Van Broekhoven et al., 2007) was applied because it incorporates the specific characteristics of fuzzy classifiers with an ordered set of classes and can deal with the fuzzy outputs of these models. Specifically, several performance criteria have been developed to evaluate and train presence-absence models, but most of these criteria are based on the confusion matrix which requires a threshold to distinguish between present and absent predictions. Since these criteria cannot deal with the fuzzy output of a fuzzy classifier, valuable information may be lost by transferring this fuzzy output to the crisp output which is needed to generate the confusion matrix. Performance measures which are derived from the confusion matrix, for instance, are not sensitive to the position of the classes where the wrong classification occurs (Van Broekhoven et al., 2007).

Therefore, Van Broekhoven et al. (2007) introduced the *AD*, which returns the average deviation between the position of the output class obtained with the model and the position of the output class stored in the training set. The *AD* varies from 0 to *n*-1 and is calculated as follows:

$$AD = \frac{1}{N} \cdot \sum_{j=1}^{N} \sum_{i=1}^{n-1} \left| \sum_{k=1}^{i} A_k(y_{model,j}) - \sum_{k=1}^{i} A_k(y_{data,j}) \right| \tag{8.1}$$

with N the number of data points, n the number of output classes, $A_k(y_{data,j})$ the membership degree of the j^{th} observed output to the k^{th} output class and $A_k(y_{model,j})$ the membership degree of the j^{th} model output to the k^{th} output class.

Although the *AD* incorporates the fuzzy characteristics of fuzzy habitat models, it depends on the prevalence of the training set and it does not consider the ecological difference between overestimation and underestimation of the observations by the model. Therefore, in this chapter, the *aAD* is introduced. This performance criterion includes a parameter α which ranges between 0 and 1 and allows stimulation of overprediction or underprediction, depending on the prevalence of the training set. To define the *aAD*, the cumulative deviation $D_{i,j}$ between the position of the class i obtained with the model and the position of the observed class i is described as

$$D_{i,j} = \sum_{k=1}^{i} A_k(y_{model,j}) - \sum_{k=1}^{i} A_k(y_{data,j}) \,. \tag{8.2}$$

The *aAD* is then defined as:

$$aAD = \frac{1}{N} \cdot \sum_{j=1}^{N} \sum_{i=1}^{n-1} \left(\frac{\left|D_{i,j}\right| + D_{i,j}}{2} + \alpha \cdot \frac{\left|D_{i,j}\right| - D_{i,j}}{2} \right). \tag{8.3}$$

Given a set of ordered output classes, the model shows overprediction if $D_{i,j}$ is negative, whereas the model underestimates the observed outputs if $D_{i,j}$ is positive. Note that the *AD* equals $\frac{1}{N} \cdot \sum_{j=1}^{N} \sum_{i=1}^{n-1} \left|D_{i,j}\right|$. In contrast to the performance measures which are based on the confusion matrix, *AD* and *aAD* are zero if the model output equals the reference output and increase with increasing distance between the reference output and the model output..

The difference between *AD*, *aAD* and confusion matrix-based performance measures such as *CCI* is illustrated in

Table 8.2. Performance measures which are based on the confusion matrix do not differentiate between situation a and b, although model prediction a is significantly better than model prediction b. The Average Deviation is not differentiating between situations c, d and e, but the *aAD* can distinguish between these three situations and will evaluate situation c as the worst situation, while situation d will be better than e depending on the value of *α*. The parameter *α* allows modellers to stimulate either overprediction or underprediction, which is clearly illustrated in situations f and g. If the model is completely underpredicting, the *aAD*

equals its maximal value 1, whereas the *aAD* equals *α* if the model is completely overpredicting. If, in this situation, *α* is lower than 1, overprediction is stimulated. Consequently, the lower the value of *α*, the more the model is trained to overpredict the observed data.

Table 8.2. Seven fuzzy classification examples and their corresponding performances expressed by *CCI*, *AD* and *aAD*. For each fuzzy set A_i of the output variable, the deviation $D_{i,j}$ is calculated based on the membership degrees of the output value to the membership functions of the fuzzy sets.

		y_{data}			y_{model}							
	A_1	A_2	A_3	A_1	A_2	A_3	$D_{1,j}$	$D_{2,j}$	$D_{3,j}$	*CCI*	*AD*	*aAD*
a	0	0.8	0.2	0	0.8	0.2	0	0	0	1	0	0
b	0	0.8	0.2	0.3	0.4	0.3	0.3	-0.1	0	1	0.4	$0.3 + 0.1\alpha$
c	0	0.8	0.2	0.6	0.2	0.2	0.6	0	0	0	0.6	0.6
d	0	0.8	0.2	0	0.2	0.8	0	-0.6	0	0	0.6	0.6α
e	0	0.8	0.2	0.4	0.2	0.4	0.4	-0.2	0	0	0.6	$0.4 + 0.2\alpha$
f	0	1	0	1	0	0	1	0	0	0	1	1
g	0	1	0	0	0	1	0	-1	0	0	1	α

Both *AD* and *aAD* can be applied to present-absent model outputs and thus to the confusion matrix. This is illustrated in Table 8.3, in which a fuzzy model output consisting of an absent and a present class is converted to a crisp output based on a threshold of 0.5.

Table 8.3. Application of the *AD* and *aAD* to presence-absence predictions and thus to the confusion matrix (0 = absent, whereas 1 = present).

Observed data	Fuzzy observed data		Crisp model output	Fuzzy model output		$\sum_{i=1}^{n-1}\left\|D_{i,j}\right\|$	$\sum_{i=1}^{n-1}\left(\frac{\left\|D_{i,j}\right\|+D_{i,j}}{2}+\alpha\cdot\frac{\left\|D_{i,j}\right\|-D_{i,j}}{2}\right)$
	Absent	Present		Absent	Present		
0	1	0	1	0	1	1	1
1	0	1	0	1	0	-1	α
0	1	0	0	1	0	0	0
1	0	1	1	0	1	0	0
						AD = 0.5	$AD = 0.25 + \alpha / 4$

Consequently, it can be shown that, for presence-absence predictions with crisp outputs,

$$AD = \frac{c+b}{N} = 1 - CCI \quad , \tag{8.4}$$

while

$$aAD = \frac{c + \alpha \cdot b}{N} \quad . \tag{8.5}$$

For presence/absence models, however, a simplified version of these criteria could be applied because only two output classes, absent (A_1) and present (A_2), are considered. In this particular case, the number of output classes, *n*, is two and thus the *AD* can be simplified as:

$$AD = \frac{1}{N} \cdot \sum_{j=1}^{N} \left| D_j \right| , \tag{8.6}$$

with

$$D_j = A_1(y_{model,j}) - A_1(y_{data,j}) \,. \tag{8.7}$$

Here $A_1(y_{model,j})$ is the normalised degree of fulfilment of the crisp output class ‘absent’ corresponding to the jth data point, and $A_1(y_{data,j})$ is one in case of absence in the dataset, zero in case of presence, for the same data point. The *aAD* is then defined as:

$$aAD = \frac{1}{N} \cdot \sum_{j=1}^{N} \left(\frac{\left|D_j\right| + D_j}{2} + \alpha \cdot \frac{\left|D_j\right| - D_j}{2} \right). \tag{8.8}$$

Consequently, if D_j is positive, the model underestimates the jth observation and the corresponding term in *aAD* is simply D_j. However, if D_j is negative, the model overestimates the jth observation and the corresponding term in *aAD* is given by $\alpha \cdot |D_j|$, a fraction of $|D_j|$, and therefore contributing less to *aAD*.

8.2.4 Training sets and scenarios

To analyse the dependency of the α parameter on the prevalence of the training set, present and absent instances of the original dataset were selected randomly to create 21 new datasets with a prevalence varying from 0 to 1 in steps of 0.05. For example, if the prevalence of the new dataset was lower than the prevalence of the original dataset (0.203 = 4579/22510), fewer present instances were selected than the number of present instances included in the original dataset, whereas all absent instances from the original dataset were selected. This resulted in a new dataset with a lower prevalence than the original dataset.

To assess the consistency of the relation between α and the training set prevalence, the aforementioned procedure was repeated with ten artificial datasets. These datasets were created based on the distribution of the environmental conditions in the original dataset because not every environmental condition occurred equally in the studied stretch. Therefore, the distribution of the original data points over the 24 environmental conditions described by the fuzzy rules was calculated. Specifically, the fuzzy sets were turned into crisp ones by assuming that an input value belongs not to a set if its membership degree to this set is < 0.5 and then each data point could be assigned to one environmental condition. Artificial datasets with the same distribution of environmental conditions as the original dataset were created by randomly generating data points which could be assigned to a specific environmental condition. The habitat suitability for spawning grayling was linked to these environmental conditions by a fixed fuzzy rule base R. Consequently, the training procedures based on different performance criteria can be compared better based on these artificial data than based on real data because the true model is known. The original dataset and the 10 artificial datasets are referred to as the initial datasets.

For each of the 11 initial datasets, 21 new datasets were created with a prevalence ranging from 0 to 1 according to the aforementioned procedure. On each of these new datasets, ten-fold cross-validation was applied to estimate the robustness of the optimisation results. The folds were constructed by randomising the new dataset and assigning each data point to one fold without replacement. The species prevalence (i.e. the frequency of occurrence) was constant for all ten folds and equal to the prevalence of the new dataset. The habitat suitability models were trained based on 25 different training scenarios (Table 8.6). The models were trained based on the *aAD* in the first 21 scenarios, with α varying from 0 (scenario 1) to 1 (scenario 21) in steps of 0.05. In the last four scenarios, the models were trained based on *CCI* (scenario 22), *Kappa* (scenario 23), the *TSS* (scenario 24) or the *AD* (scenario 25). Consequently, 57750 (11 x 21 x 10 x 25) different simulations were performed on a Linux cluster containing 14 nodes, containing 72 Dual-Core Intel® Xeon® CPU 3.00 GHz processors in total, 1 Gb of RAM and running a 2.6.5 kernel.

8.3 Results

Based on the original dataset, 21 new sets with different prevalences were created and rule base training on these sets, based on the *aAD* with varying *α*, resulted in different final rule bases at a specific training set prevalence (Fig. 8.1). Although general agreement among these rule bases was observed, some rule bases showed slight differences at different prevalence values. This result could be explained by the presence of 'core rules' and 'ghost rules' in the trained rule base. Specifically, the fuzzy rules in a trained rule base represented each possible combination of input variable sets, but some rules may describe a combination of input variable sets which was absent in the studied stretch. These rules are referred to as ghost rules, whereas core rules describe a environmental condition which is present in the studied stretch. In contrast to the core rule consequents, the ghost rule consequents could not be trained based on the available data and could thus randomly take any linguistic value of the output variable. However, these ghost rules did not affect the model predictions because they did not represent the studied stretch. Consequently, rule base training in this chapter sometimes resulted in rule bases with different ghost rule consequents, but equal core rule consequents and thus equal model predictions. The trained rule bases in this chapter could therefore be joined in 11 clusters (Fig. 8.1) based on the consequents of their core rules.

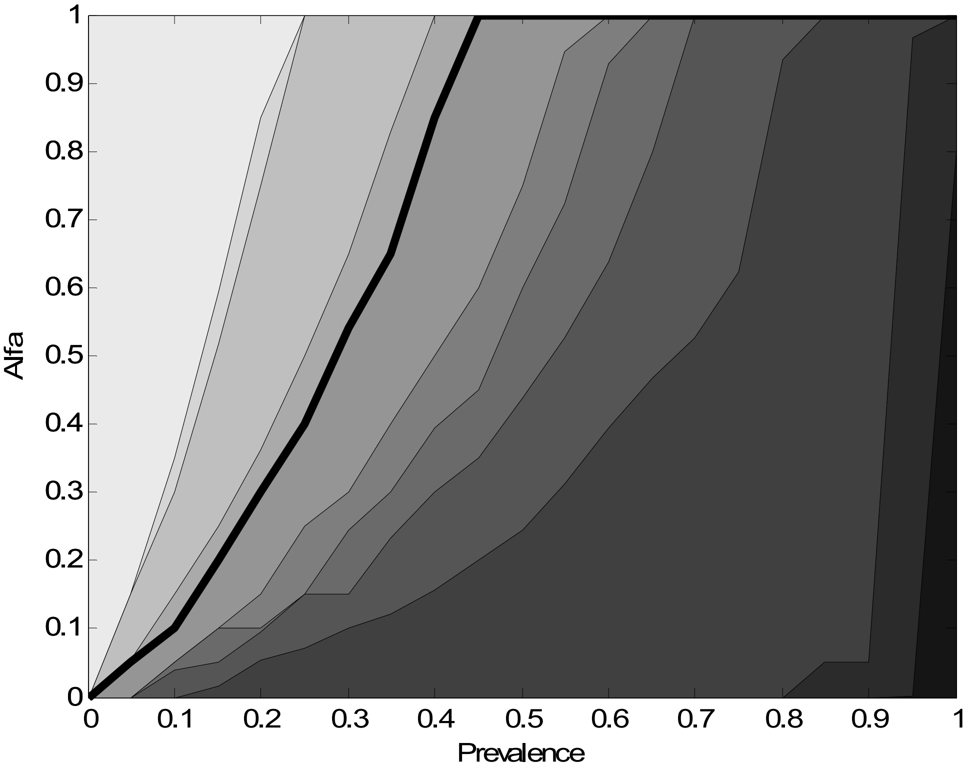


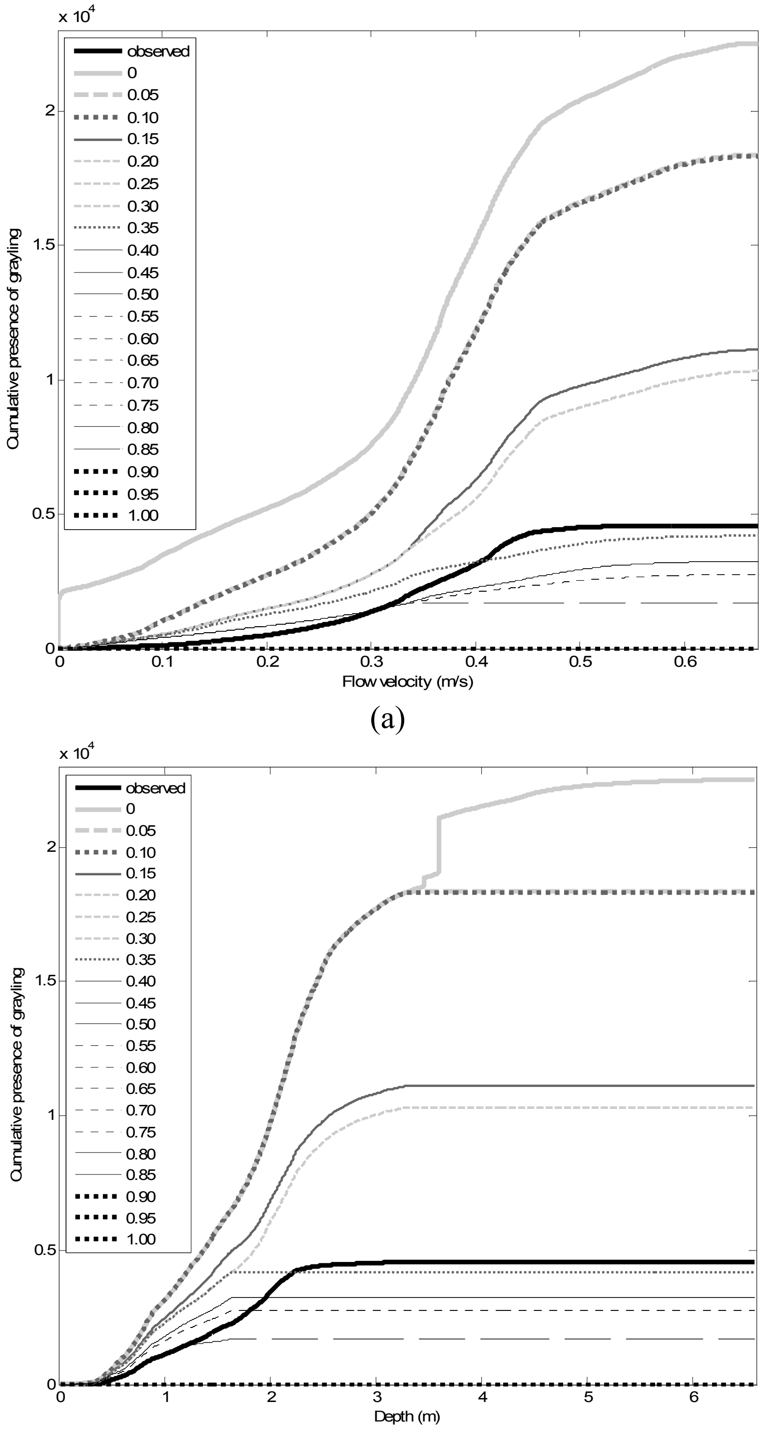
Fig. 8.1. The 11 different rule base clusters obtained from rule base training based on *aAD* with varying *α*. Training was performed based on 21 training sets with a prevalence ranging between 0 and 1 in 0.05 steps, which were derived from the original dataset. Areas with the same shade represent the same rule base cluster, while the brightness of an area reflects the number of 'absent' consequents in the core rules of the rule base. The light area (left top) contains the rule base with no 'present' consequents, whereas the darkest area (right down) represents the rule base with no 'absent' consequents. The left boundary of each area is referred to as the *α* curve of this cluster and connects the maximum values of *α* at which the rule base of this cluster was obtained. These values of *α* were averaged over the ten training folds at each prevalence. The bold black boundary between two clusters indicates the maximum *α* at which the first rule base is obtained that overestimates the observations.

In Fig. 8.1, the left boundary of each cluster is referred to as an α curve and connects the maximum values of α at which a rule base of this cluster was obtained. The results were consistent across the different folds, with the standard deviation of the values of α of the α curves averaging 0.006 and ranging between 0 and 0.066. Not all 11 rule base clusters were obtained at all prevalence values, while most of the clusters were obtained at intermediate prevalence values (approximately between $P = 0.1$ and $P = 0.7$).

The number of 'present' relevant rule consequents in the final rule base was negatively correlated with the value of the α parameter at each prevalence (Fig. 8.1). Consequently, at a fixed training prevalence, α represents the likelihood that a rule base is obtained which underestimates the observations: the lower α, the higher the likelihood that an overestimating rule base is found. The fifth cluster (starting from the top left corner in Fig. 8.1) contains the first rule base which overestimates the observations (Fig. 8.2). The upper boundary of this cluster (the solid black line in Fig. 8.1) represents the maximum values of α at which the overestimating rule base is obtained. If at a specific prevalence, α is lower than or equal to the corresponding α on this line, the final rule base will overestimate the observations. The solid black line (Fig. 8.1) represents the transition from an underestimating rule base to an overestimating one. Consequently, the rule base which approximates the observations most accurately, could be found at values of α ranging between the solid black line and the upper boundary of the first rule base cluster above this line.

This is also reflected in the cumulative plots of the predictions of the habitat suitability of spawning grayling by the different rule bases for the input variables flow velocity and depth (Fig. 8.2 a-b). Not only do these plots provide an indication of the ecological relevance of the optimised rule bases, they also show the similarity between the different rule bases. Although the rule base obtained with $\alpha = 0.35$ approximates the total number of observed occurrences relatively accurately, this rule base is still underestimating the occurrence of spawning grayling. Since the rule base obtained with $\alpha = 0.30$ does overestimate the grayling presence substantially, the most accurate rule base might be found at values of α between 0.35 and 0.30.

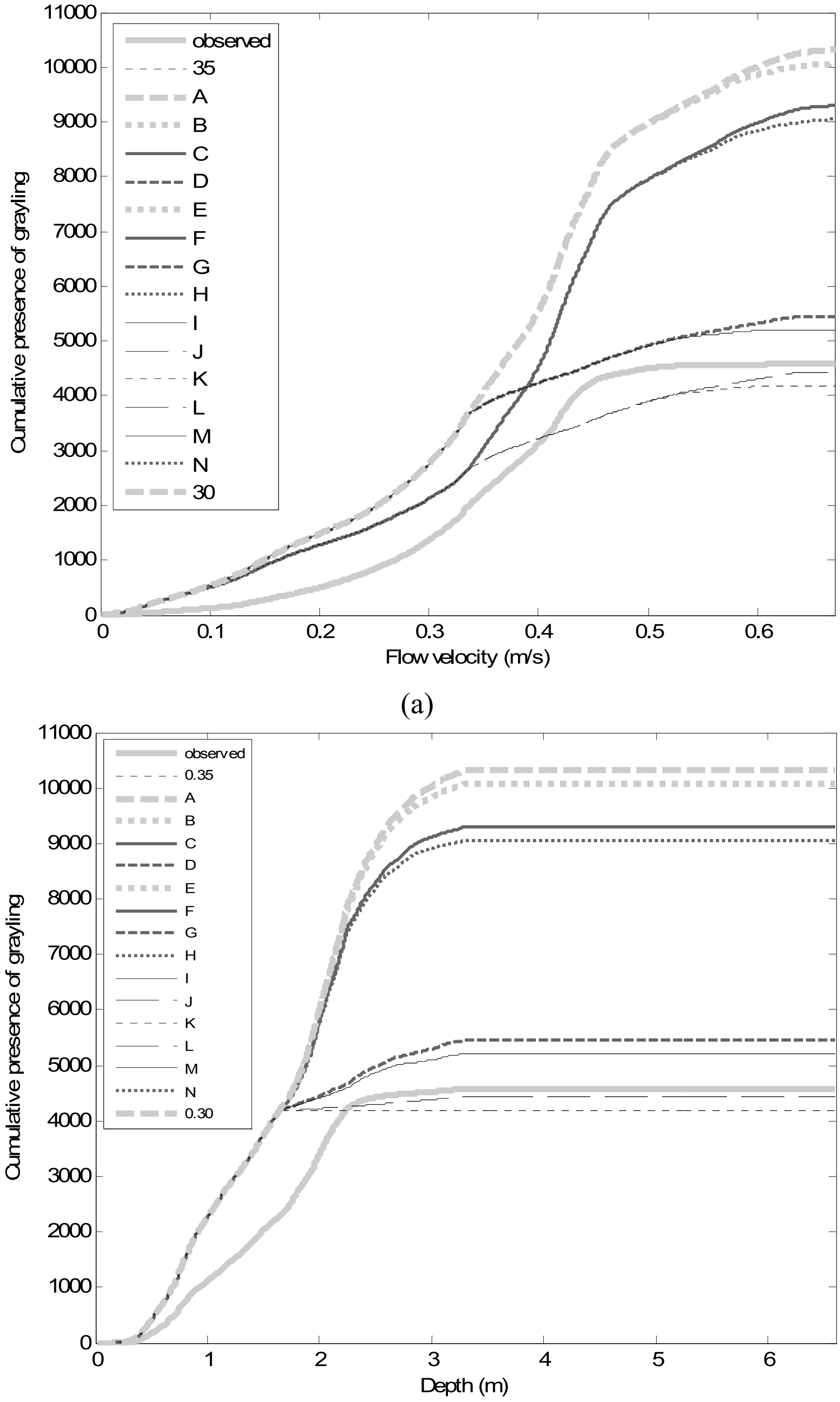
Further analysis of both rule bases revealed that four consequents differed between both rule bases (Table 8.5). Consequently, the transition between both rule bases consists of 14 different rule bases, one of which may be the most accurate rule base (Table 8.5). The cumulative predictions of these rule bases (Fig. 8.3) show that 12 rule bases are approximating the observations more accurately (all except A and K). However, the rule bases which approximate the observations the best (D, G, I, J, K, L, M), show some substantial differences with the observations. For flow velocity, for instance, the gradient of the observed curve was the highest between 0.27 $m.s^{-1}$ and 0.43 $m.s^{-1}$, whereas for these 7 rule bases, this gradient is the highest between 0.27 $m.s^{-1}$ and 0.33 $m.s^{-1}$. Similar results were found for depth, where most spawning grayling was observed between 1.9 and 2.3 m, but these 7 rule bases predict grayling to occur the most between 1 and 1.7 m. In contrast to these 7 rule bases, the shape of the cumulative predictions of the other rule bases (A, B, C, E, F, H, N) is much more similar to the shape of the cumulative observations. This indicates that, even though these rule bases are substantially overestimating the observations, they may be ecologically more relevant than the first 7 rule bases. Similar results were obtained for the other two input variables.



(a)

(b)

Fig. 8.2. Cumulative plots of the observations and of the rule base predictions obtained after training based on *aAD* with *α* varying between 0 and 1 in 0.05 steps. Predictions were averaged over the ten folds and cumulated according to their corresponding flow velocity (a) or depth (b) values. Training was performed on the original dataset with a prevalence of 0.20. Values of *α* resulting in the same core rule base, are indicated by the same line type. Cumulative predictions based on the other two input variables showed similar results but are not displayed since these plots are harder to interpret due to the ordinal nature of these variables.



(a)

(b)

Fig. 8.3. Cumulative plots of the observations and of the predictions of the 14 rule bases which are situated between the rule base obtained with $\alpha = 0.30$ and the one obtained with $\alpha = 0.35$ (Table 8.5). Predictions were averaged over the ten folds and cumulated according to their corresponding flow velocity (a) or depth (b) values. Training was performed on the original dataset with a prevalence of 0.20. Rule bases with the same core rule base are indicated by the same line type. Cumulative predictions based on the other two input variables showed similar results but were not shown since these plots are harder to interpret due to the ordinal character of these variables.

Table 8.4. The 4 rule consequents (*a*, *b*, *c* and *d*) which distinguish the rule base obtained with $\alpha = 0.30$ from the one obtained with $\alpha = 0.35$ (A = absent, P = present).

				% fine gravel			
				low		high	
				% medium-sized gravel		**% medium-sized gravel**	
				low	high	low	high
Flow velocity	low	**Depth**	low	P	P	P	A
			moderate	*a*	A	*c*	A
			high	A	A	A	A
	high	**Depth**	low	P	P	P	A
			moderate	*b*	A	*d*	A
			high	A	A	A	A

Table 8.5. The 14 rule bases which are situated between the rule base obtained with $\alpha = 0.30$ and the one obtained with $\alpha = 0.35$. The four consequents that differ between these two rule bases were labelled *a* to *d*. The values of these consequents are given for the 14 rule bases, while all other consequents remained constant (A = absent; P = present).

	Consequent values			
Rule base	*a*	*b*	*c*	*d*
$\alpha = 0.30$	P	P	P	P
A	A	P	P	P
B	P	A	P	P
C	P	P	A	P
D	P	P	P	A
E	A	A	P	P
F	A	P	A	P
G	A	P	P	A
H	P	A	A	P
I	P	A	P	A
J	P	P	A	A
K	P	A	A	A
L	A	P	A	A
M	A	A	P	A
N	A	A	A	P
$\alpha = 0.35$	A	A	A	A

Comparison of the training based on *aAD* with training based on four other performance criteria revealed that the latter four criteria produced rule bases which underestimated the observations (Fig. 8.4). However, the shape of the cumulative predictions obtained by training based on *Kappa* and on the *TSS* was more similar to the shape of the cumulative observations than it was for the underestimating rule bases derived based on *CCI* and *AD*. This may indicate that the rule bases obtained from *Kappa* and *TSS* training may be ecologically more relevant than those obtained after training based on *aAD* with a value of α that leads to underestimating rule bases (Fig. 8.2). Nevertheless, the shape of the cumulative predictions obtained from training based on *aAD* with a value of α of 0.30 (Fig. 8.4) still corresponds the most to the observations. Moreover, other performance criteria indicate that model performance is acceptable for this training scenario (Table 8.6).

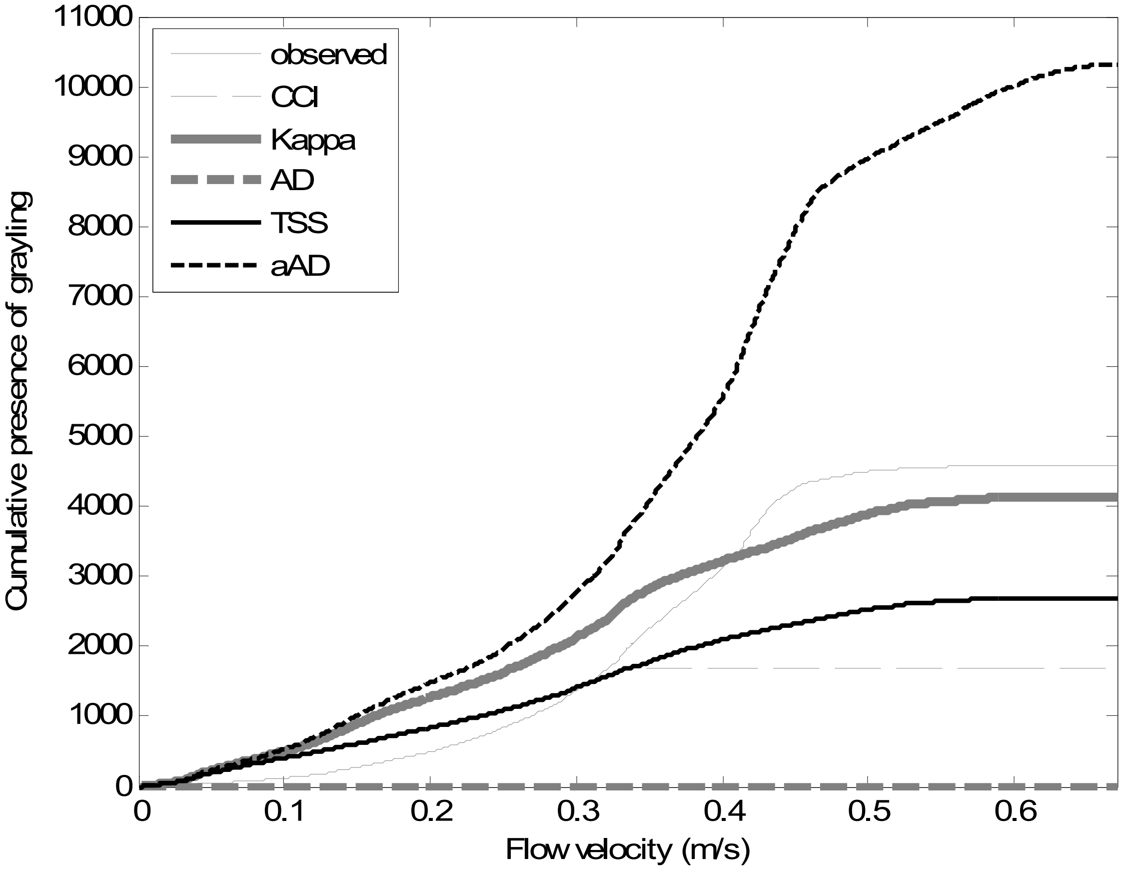


Fig. 8.4. The cumulative predictions of grayling for the variable flow velocity obtained after training based on the percentage of correctly classified instances (*CCI*), *Kappa*, the average deviation (*AD*), the true skill statistic (*TSS*), and the adjusted average deviation (*aAD*) with α = 0.30. Training was performed on the original dataset.

Table 8.6. The values of 4 performance criteria (the percentage of correctly classified instances *CCI*; Cohen's *Kappa*; the true skill statistic *TSS* and the average deviation *AD*), calculated from the predictions which were obtained based on the 25 training scenarios. The scenarios are indicated by the value of α if *aAD* is applied for model training and by their performance criterion if another performance criterion is applied for model training. The value of the *aAD* is not given because it depends on α and thus cannot be compared across different training scenarios.

Scenario description	*CCI*	*Kappa*	*TSS*	*AD*
0	0.1999	0	0	0.8001
0.05	0.3851	0.1075	0.2312	0.6179
0.1	0.3860	0.1074	0.2311	0.6142
0.15	0.6421	0.2771	0.3337	0.4008
0.2	0.6667	0.2982	0.3492	0.3702
0.25	0.6667	0.2982	0.3492	0.3702
0.3	0.6667	0.2982	0.3492	0.3702
0.35	0.7854	0.3123	0.3044	0.2386
0.4	0.7961	0.2893	0.2591	0.2270
0.45	0.7961	0.2893	0.2591	0.2270
0.5	0.7961	0.2893	0.2591	0.2270
0.55	0.8010	0.2732	0.2349	0.2146
0.6	0.8010	0.2732	0.2349	0.2146
0.65	0.8010	0.2732	0.2349	0.2146
0.7	0.8010	0.2732	0.2349	0.2146
0.75	0.8010	0.2732	0.2349	0.2146
0.8	0.8037	0.1983	0.1539	0.2080
0.85	0.8037	0.1983	0.1539	0.2080
0.9	0.8001	0	0	0.1999
0.95	0.8001	0	0	0.1999
1	0.8001	0	0	0.1999
Kappa	0.7903	0.3240	0.3063	0.2632
CCI	0.8041	0.2027	0.1495	0.2133
AD	0.8001	0	0	0.1999
TSS	0.8037	0.2769	0.3580	0.2135

Similar results were obtained for the 10 artificial datasets as for the original dataset. Rule base optimisation based on *CCI*, *Kappa*, *TSS* and *AD* resulted in the true model after model training, whereas optimisation based on the *aAD* led to different final models at different values of α. This reflected the results obtained from training on the original rule base. At each prevalence, the true model was obtained at the same value of α for all artificial datasets. Fig. 8.5 compares these values of α with the values of α at which the first rule base is obtained that overpredicts the observations. At each prevalence value, the latter values of α are lower than or equal to the first values of α. This indicates that overprediction should be stimulated more for the original dataset, which contains data noise, than for the artificial datasets.

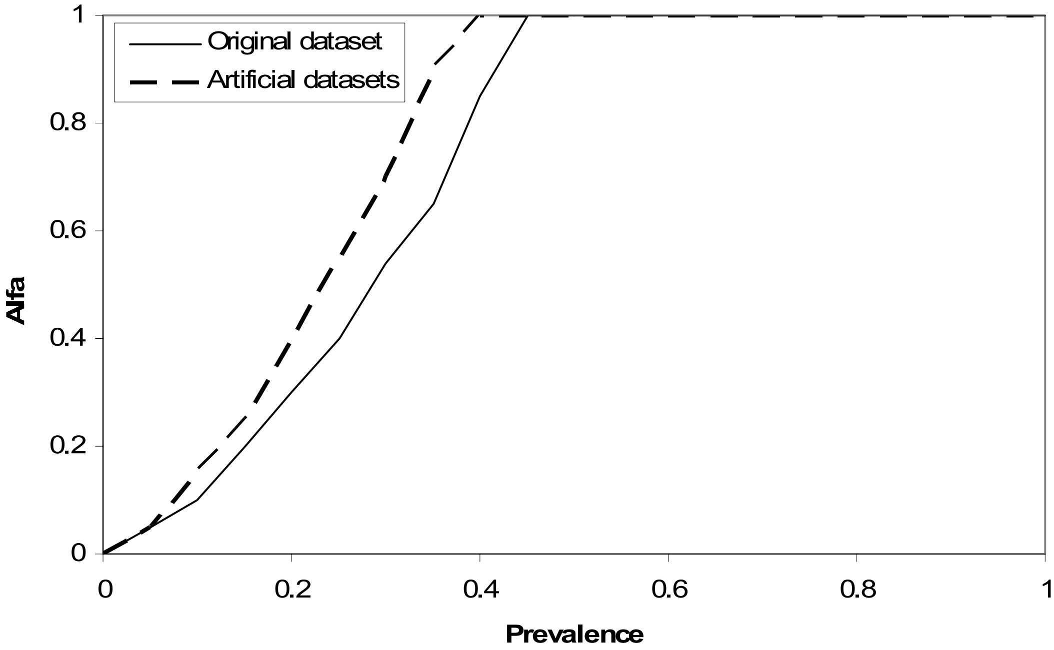


Fig. 8.5. Comparison of the values of α at which the true model is obtained after training on the artificial datasets and the values of α at which the first rule base is obtained that overpredicts the observations. The latter values of α are obtained from training based on the original dataset.

8.4 Discussion

Although an optimal parameter value could be found by applying sensitivity analysis, a more important problem with these flexible performance criteria could be the difficulty to say which models are better (Glas et al., 2003; Vaughan and Ormerod, 2005). The results presented in this chapter showed that comparison of the shapes of the cumulative prediction curves may provide an indication of the ecological relevance of the different optimised models. Specifically, observations may be more appropriate than expert knowledge to evaluate model prevalence because differences in species' dispersal patterns and associated gene flow may lead to subtle variations in habitat preferences of some species due to local adaptations (Holt, 2003). Even in the absence of genetically driven differences in habitat use, species could express different realised niches (Hutchinson, 1957) as a result of spatial variation in predators, competitors or other biotic factors (Hutchinson, 1957; Osborne and Suarez-Seoane, 2002; Holt, 2003; Peterson and Holt, 2003; Hernandez et al., 2006; McPherson and Jetz, 2007). Consequently, conservationists should distinguish between models which reliably or less reliably predict species distribution (McPherson and Jetz, 2007). Maggini et al. (2006) show that applying a training parameter or threshold which minimises the difference between over- and underprediction may be a suitable rule of thumb. Applied to the results of this chapter, this suggests that the optimal rule base could be obtained by applying a value of α between 0.35 and 0.30.

However, sometimes specific knowledge on model selection is available, for example if omission errors should be considered more 'dangerous' in the application than omission errors. The parameter α in the *aAD* allows to incorporate a certain degree of this knowledge in the data-driven model development process because this parameter allows modellers to focus on underprediction or overprediction during model training. If, for instance, the modellers know that the training set has an artificially low prevalence, they could choose to stimulate overprediction because such low prevalence could be the result of inefficient monitoring, missing data, ecological effects such as competition and predation, or temporal effects such as seasonality and migration barriers. Another, but more intensive option for α setting is selecting certain field observations which should be predicted as present or absent by the model, based on ecological expert knowledge. As such, the modeller could assess the consistency between the available ecological expert knowledge and the observed data.

Consequently, it may be more appropriate to apply weights in the training performance criteria than to weight absences in the training data because the first approach allows modellers to adjust the model training process to the purpose to the application. Jiménez-Valverde and Lobo (2006) argue that unbalanced species distribution data are not such a problem from a statistical point of view, and that the effects unbalanced prevalence should not be confused with those of low-quality data affected by false absences for example. Consequently, it makes more sense to focus on the relation between omission and commission errors during model training than to adjust training data to a prevalence of 0.5. Further research should reveal whether the same model predictions are obtained after threshold optimisation during model evaluation as the predictions which are obtained by flexible training with a threshold-dependent performance criterion. Another aspect worth investigating is whether weighting absences to ensure a prevalence of 0.5 (Maggini et al., 2006) results in the same predictions as the predictions obtained after flexible training. Given the results of Jiménez-Valverde and Lobo (2006), it could be expected that the difference between the results of both approaches is negatively correlated with the quality of the training data.

The results of this chapter show that the four common performance criteria may be less appropriate for model training. Not only do the results of all four model training scenarios based on the four criteria appear to depend on the prevalence of the training set, the shapes of the cumulative prediction curves also differ substantially from the shape of the cumulative observations (Fig. 8.4). Moreover, none of these criteria allows comparison of different training results to select the optimal model. However, model developers should be able to carefully choose an appropriate training performance criterion which reflects the ecological model purpose (Segurado and Araújo, 2004).

Surprisingly, the results of model training based on *Kappa* differ substantially from those from the previous study in Chapter 7. In Chapter 7, rule base training based on *Kappa* resulted in a rule base that overestimated the observations, whereas in this chapter the final rule base obtained after *Kappa* training underestimated the observations. A possible explanation may be the different fuzzy sets which were applied in both studies. Specifically, the fuzzy sets in Chapter 7 were derived from expert knowledge, whereas the sets in this chapter were obtained from the entropy-based approach. This led to a different distribution of the training samples over the environmental conditions which were represented by the fuzzy sets. Consequently, the distribution of the present and absent instances over the environmental conditions also differed between both chapters. This may have resulted in a different focus of *Kappa* during rule base optimisation as was described in Chapter 6. Specifically, Chapter 6 showed that the focus of *Kappa* depends on the relations between the elements of the confusion matrix, and these relations may be affected by the distribution of the present and absent instances over the input space. This assumption is supported by the similarity of the results obtained from rule base optimisation based on *CCI* in Chapter 7 and 8. Chapter 6 showed indeed that the focus of *CCI* only depends on the prevalence of the training set and thus the results in Chapter 7 and 8 based on *CCI* optimisation should agree because the prevalence is equal in both chapters.

Based on the aforementioned distinction between omission and commission errors and on the similarity between the shape of the cumulative predictions curve with that of the cumulative observation curve, the optimal rule bases appear to be C and H (Fig. 8.3, Table 8.5). The predictions of these rule bases are in line with those obtained from the same dataset in Chapter 7 and thus may show substantial ecological relevance. Further analysis also indicated that the training set prevalence also had an effect on the results of the *Kappa* training (Fig. 8.4). The similar results for *CCI*, *AD* and *TSS* suggest that these criteria may also be less appropriate for model training. However, Table 8.6 confirms the value of *TSS* and *Kappa* for the evaluation of the performance of the final model. Specifically, the rule bases from the first 21 training scenarios which showed acceptable ecological relevance (α between 0.15 and 0.35), were also characterised by relatively high *TSS* and *Kappa* values. Both accuracy-based measures, *CCI* and the *AD*, appeared to be less suitable for model evaluation because their values were correlated with the value of α (R = 0.58 and 0.65 respectively).

This chapter confirms the results of Chapter 7 by showing that the performance criteria which are applied for model training may significantly affect the results of the training process. Consequently, these findings suggest that model developers should use training scenarios based on different performance criteria or on an adjustable parametrised criterion. As such, they should carefully select an appropriate training performance criterion or parameter which reflects the ecological model purpose. Although this approach is illustrated in this chapter by applying the adjusted average deviation to train a fuzzy habitat suitability for spawning grayling, the presented methods could be applied to any ecosystem or species. Therefore, this chapter may contribute to improve the reliability of ecological models in general and thus provide valuable insights for ecosystem management.

CHAPTER 9

Sensitivity analysis of prevalence-adjusted training

9.1 Introduction

The previous chapter showed that prevalence-adjusted training performance criteria should enable model developers to adjust the training procedure to the training set prevalence. Chapter 7 demonstrated that such criteria may provide insight in the focus of the training procedure and in the ecological relevance of the fitted models. Specifically, the model training procedure of a habitat suitability model for European grayling (*Thymallus thymallus*) in a Swiss river was adjusted to the training set prevalence by varying the parameter α of an prevalence-adjusted performance criterion, the adjusted average deviation (*aAD*).

This chapter aims to analyse the relation between the parameter α and the training set prevalence for three other species in different New Zealand river systems: the caddisflies *Aoteapsyche* spp., large brown trout *Salmo trutta* and rainbow trout *Oncorhynchus mykiss*. Like in the previous study, the *aAD* was implemented in a hill-climbing algorithm to optimise a fuzzy species distribution model for each species. Specifically, the hypotheses were tested that (1) similar relations between the parameter α and the training set prevalence would be obtained, (2) training based on the *aAD* would lead to more accurate model predictions than training based on more frequently applied performance criteria such as *CCI*, and that (3) the final fuzzy model would produce a realistic model of habitat suitability.

9.2 Material and methods

9.2.1 Study area and collected data

Benthic invertebrate data were collected by Surber sampling in six New Zealand rivers: the Mangles, Mohaka, Waingawa, Clutha (Jowett et al., 1991), Waitaki (Stark and Suren, 2003) and Tongariro (Collier, 1993) rivers. From each of the 692 samples, caddis fly larvae *Aoteapsyche* spp. were selected. *Aoteopsyche* is a net-spinning caddis fly and is typically found in rivers where the substrate is relatively stable. Larvae were found in 488 samples, which resulted in a prevalence of 0.71. Samples were collected from as wide a range of water depths, velocities and substrate types as possible in each river. Mean water column velocity (at 60 % of the depth) and water depth were measured at each sampling point with a current meter on a calibrated rod. The substrate composition was estimated visually with a modified Wenthworth particle size scale. The substrate composition was converted into a single index (*s*) by summing weighted percentages of each substrate type (Jowett et al., 1991). To allow for a fine-gravel category, the weighting values used were a slightly modified form of the original instream flow incremental methodology substrate codes (Bovee, 1982):

$s = 0.08 \,.\, \text{bedrock} + 0.07 \,.\, \text{boulder} + 0.06 \,.\, \text{cobble} + 0.05 \,.\, \text{gravel} + 0.04 \,.\, \text{fine gravel} + 0.03 \,.\, \text{sand}.$

The habitat use by trout was measured in the Lake Wanaka outlet of the Clutha river in February 2005. The Clutha river is one of New Zealand's largest rivers and supports a very high density of large brown trout *Salmo trutta* and rainbow trout *Oncorhynchus mykiss* (Teirney and Jowett, 1990). At the time of the survey, the river flow was 226 m s^{-1} (mean flow = 195 m s^{-1}), the average width was about 90 m, and the maximum depth was 7 m. Black disk visibility was 9.5 m and thus a diver could observe fish in what were considered undisturbed locations. Specifically, fish that used the entire water column for feeding would move laterally or, more often, move closer to the bottom as the diver approached. Fish were identified as brown trout or rainbow trout and as large-sized (> 40 cm) or medium-sized (20-40 cm) fish. This study only focused on large brown trout and rainbow trout. Large brown trout was observed less frequently than large rainbow trout, resulting in respective prevalences of 0.03 and 0.11. While the diver located and counted fish, an accompanying boat recorded water depths and flow velocities with an acoustic Doppler current profiler (ADCP; Teledyne RD Instruments, Poway, California). Although precise locations of individual fish could not be located accurately, the river channel was straight and had few abrupt variations in depth, either longitudinally or laterally, so that ADCP measurements of depth and velocity could be assumed to apply at fish locations. For each fish location, a measurement of depth and depth-averaged velocity was randomly selected from the relevant ADCP file. Examination of the ADCP data confirmed that there was little variation in depth and velocity in the ADCP measurement, and only two measurements (1.5 %) were excluded because the depth and velocity at the fish location could not be ascertained with sufficient certainty. Substrate throughout the reach was generally a mixture of boulder, cobble and gravel; cobbles were usually the most common substrate. Substrate was not considered further because it was relatively constant and appeared to have little functional relevance for the trout. Fish locations were recorded with equal effort in all habitat types and were supplemented by some bank observations of trout in water near the shore in areas where the diver or boat team was unable to operate. Habitat availability data were collected at an average of 2.9 m intervals across 15 randomly selected cross-sections at a flow of about 170 m s^{-1}. Water surface profile modelling (RHYHABSIM; Jowett, 1996) was used to predict depths and velocities at 226 m s^{-1}, the flow at which the habitat use data were collected. Additional cross-sectional data were collected at a flow of 226 m s^{-1} with the ADCP. The modelling approach in this chapter was identical to the procedure described in Chapter 8. However, only 5250 (21 x 10 x 25 different simulations were performed because no artificial datasets were created.

Table 9.1. Input variables recorded and the corresponding fuzzy sets of the species distribution models. The entropy, indicating the uniformity of the distribution of the values of a variable over its fuzzy sets, was calculated for all input variables.

Species	Input variable	Fuzzy set	Parameters	Entropy
Aoteapsyche spp.	Depth (m)	Shallow	(0.00,0.00,0.31,0.79)	0.74
		Moderate	(0.31,0.79,1.08,1.32)	
		Deep	(1.08,1.32,2.40,2.64)	
	Flow velocity (m/s)	Low	(0.00,0.00,0.45,1.34)	0.85
		High	(0.45,1.34,1.78,2.23)	
	Substrate index (-)	Low	(0.00,0.00,1.73,5.18)	0.30
		High	(1.73,5.18,6.90,8.63)	
Large brown trout	Depth (m)	Low	(0.00,0.00,0.89,2.31)	0.93
		Medium	(0.89,2.31,3.20,3.91)	
		High	(3.20,3.91,7.11,7.82)	
	Flow velocity (m/s)	Low	(0.00,0.00,0.62,1.85)	0.91
		High	(0.62,1.85,2.46,3.08)	
Large rainbow trout	Depth (m)	Low	(0.00,0.00,1.73,5.18)	0.86
		High	(1.73,5.18,6.91,8.63)	
	Flow velocity (m/s)	Low	(0.00,0.00,0.62,1.85)	0.91
		High	(0.62,1.85,2.46,3.08)	

9.3 Results

9.3.1 Caddis fly

Based on the original dataset, 21 new sets with different prevalences were created and rule base training on these sets, based on the aAD with varying α, resulted in different final rule bases at a specific training set prevalence (Fig. 9.1). Although general agreement among these rule bases was observed, some rule bases showed slight differences at different prevalence values. This result could be explained by the presence of ‘core rules’ and ‘ghost rules’ in the trained rule base. Specifically, the fuzzy rules in a trained rule base represented each possible combination of input variable sets, but some rules may describe a combination of input variable sets which was absent in the studied stretch. These rules are referred to as ghost rules, whereas core rules describe a environmental condition which is present in the studied stretch. In contrast to the core rule consequents, the ghost rule consequents could not be trained based on the available data and could thus randomly take any linguistic value of the output variable. However, these ghost rules did not affect the model predictions because they did not represent the studied stretch. Consequently, rule base training in this paper sometimes resulted in rule bases with different ghost rule consequents, but equal core rule consequents and thus equal model predictions. The trained rule bases in this paper could therefore be joined in 5 clusters (Fig. 9.1) based on the consequents of their core rules.

Most of the values of α which delimited the clusters were increasing monotonously with the prevalence, but at higher prevalence values some irregularities occurred (Fig. 9.1). The number of ‘present’ consequents in the final rule base was negatively correlated with the value of the α parameter at each prevalence (Fig. 9.1). Consequently, at a fixed training prevalence, α represents the likelihood that a rule base is obtained which underestimates the observations: the lower α, the higher the likelihood that an overestimating rule base is found.

The fourth cluster (starting from the top left corner in Fig. 9.1) contains the first rule base which overestimates the observations. The upper boundary of this cluster (the solid black line in Fig. 9.1) represents the maximum values of α at which an overestimating rule base is obtained. If at a specific prevalence, α is lower than or equal to the corresponding value of α on this line, the final rule base will overestimate the observations. The solid black line (Fig. 9.1) represents the transition from an underestimating rule base to an overestimating one. Consequently, the rule base which approximates the observations most accurately, could be found at values of α ranging between the solid black line and the upper boundary of the first rule base cluster above this line.

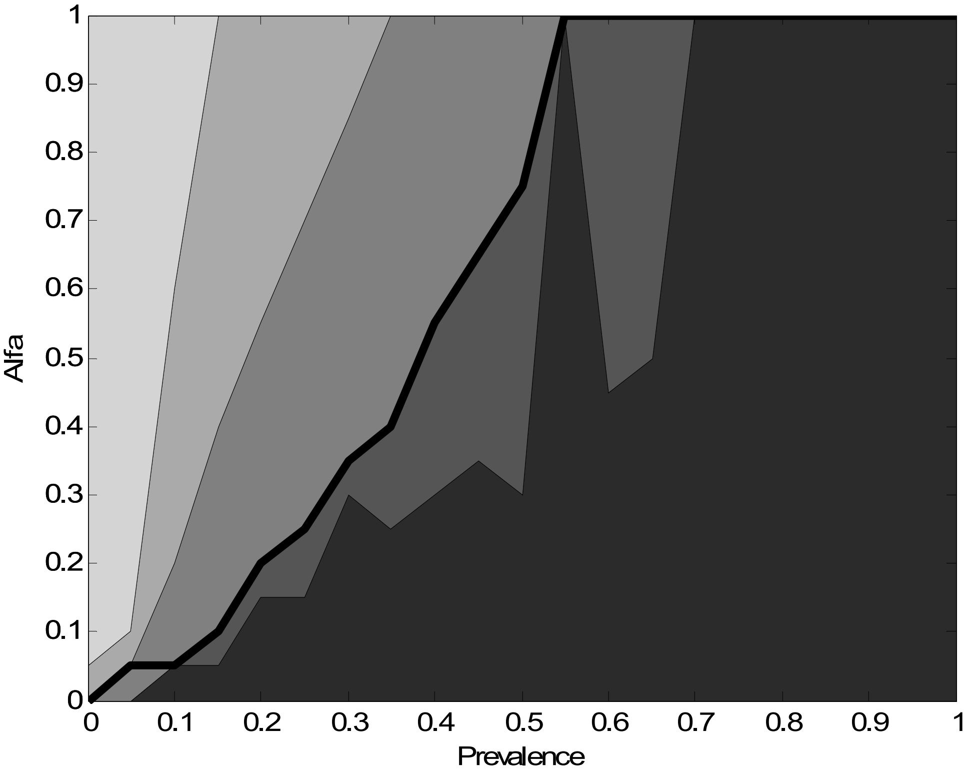


Fig. 9.1. The 5 different rule base clusters obtained from rule base training based on *aAD* with varying α, averaged over the ten training folds at each prevalence of *Aoteapsyche* spp. Training was performed based on 21 training sets with a prevalence ranging between 0 and 1 in steps of 0.05, which were derived from the original dataset. Areas with the same shade represent the same rule base cluster, while the brightness of an area reflects the number of ‘absent’ consequents in the rule base. The light area (left top) contains the rule base with no ‘present’ consequents, whereas the darkest area (right down) represents the rule base with no ‘absent’ consequents. The left boundary of each area connects the maximum values of α at which the rule base of this area was obtained. The black line indicates the maximum values of α at which the first rule base was obtained that overestimated the observations.

The results were identical across the different folds, but not all 5 rule base clusters were obtained at each prevalence value. Consequently, plotting the cumulative predicted presence of the original dataset (with prevalence 0.71), revealed only two possible solutions, which is illustrated for the flow velocity in Fig. 9.2. The effect of the α parameter on the training results therefore becomes more clear by plotting the cumulative predicted presences at a lower prevalence value (Fig. 9.3). The cumulative plots provide an indication of the ecological relevance of the optimised rule bases and show the similarity between the different rule bases. Since the rule bases obtained with α = 0.25 and α = 0.20 respectively underestimate or overestimate the caddis fly presence substantially, the most accurate rule base may be found at values of α between 0.25 and 0.20 (Fig. 9.3).

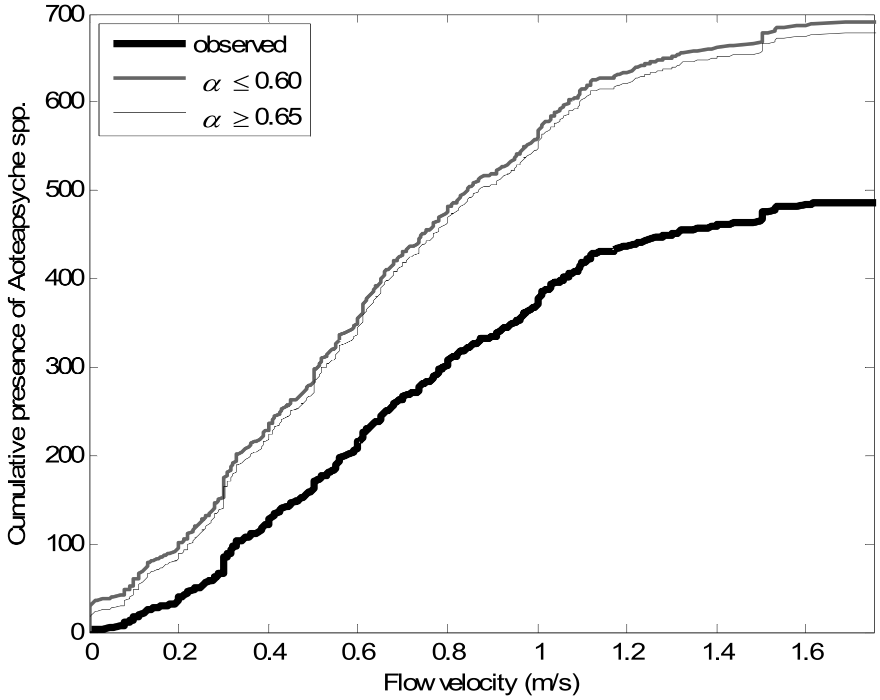
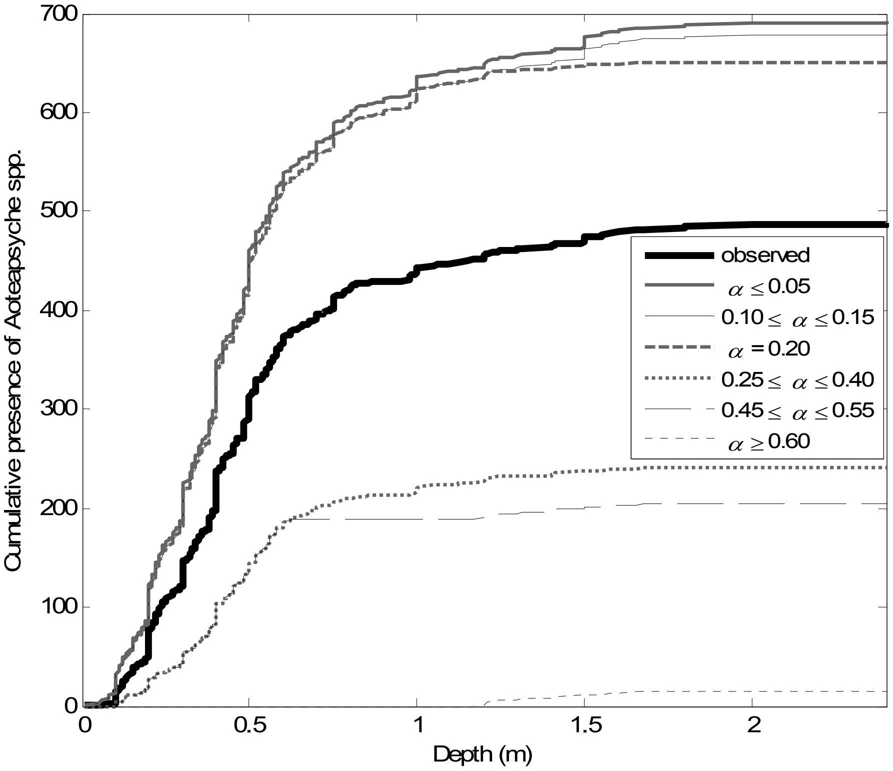
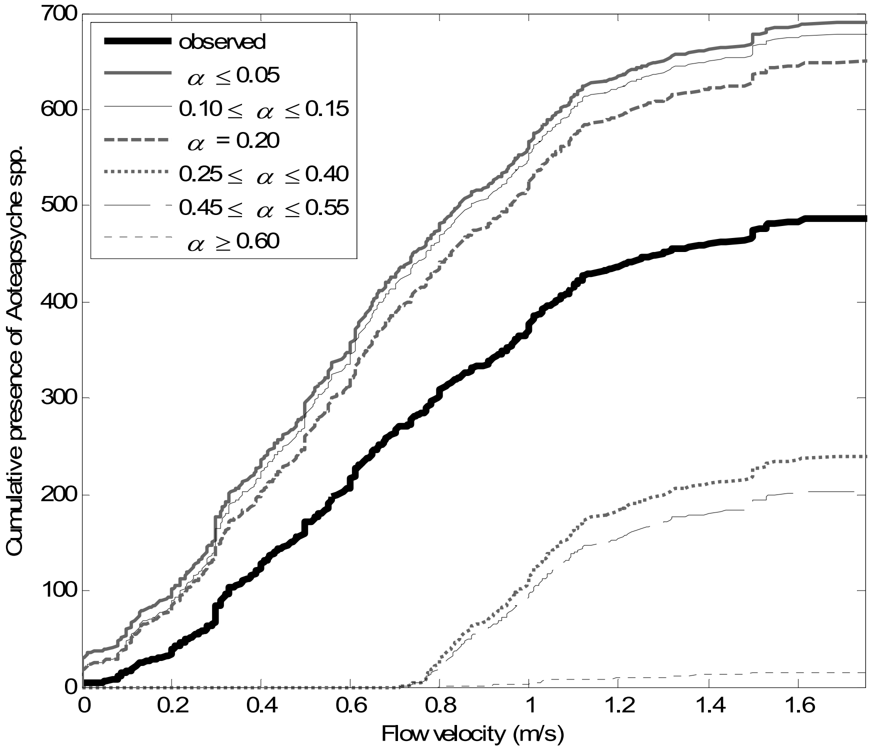


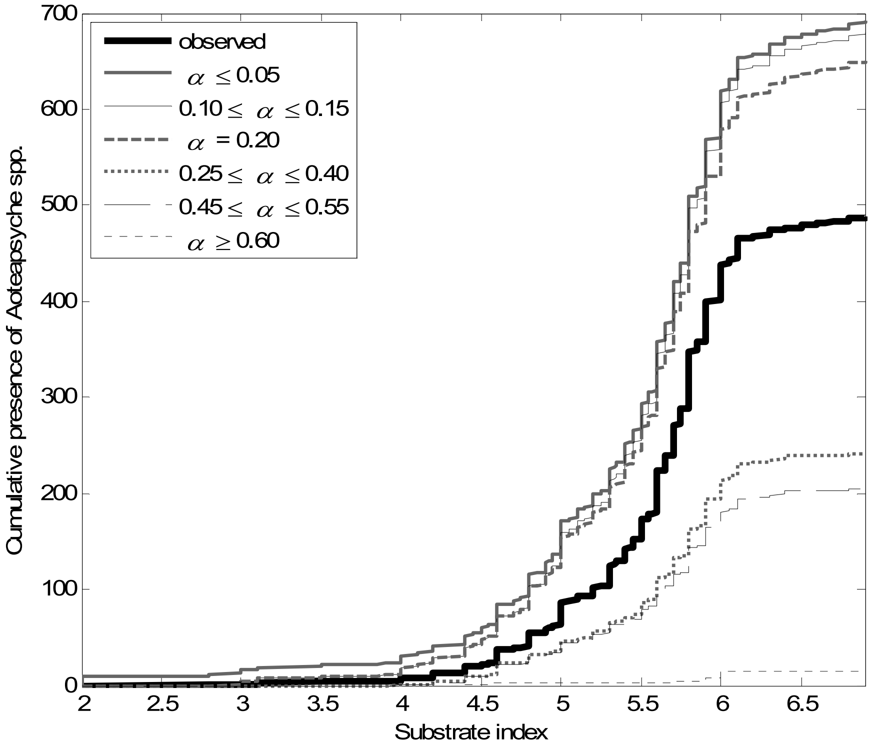
Fig. 9.2. Cumulative plots of the observations and of the rule base predictions of *Aoteapsyche spp.* obtained after training based on *aAD* with values of *α* varying between 0 and 1 in steps of 0.05. Predictions were averaged over the ten folds and cumulated according to their corresponding flow velocity values. Training was performed on the original dataset with a prevalence of 0.71. Values of *α* resulting in the same core rule base, are indicated by the same line type. The lines show the predictions based on the rule bases obtained after training with *α* varying between 0 and 0.60 (bold grey line) and between 0.65 and 1 (thin black line).



(a)

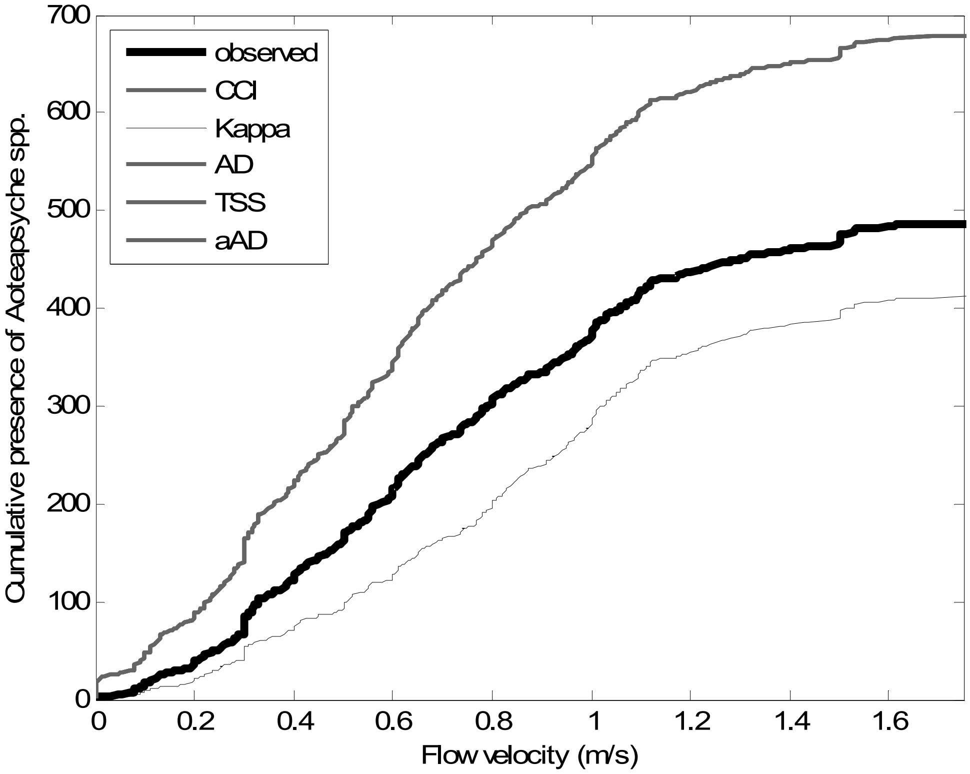


(b)

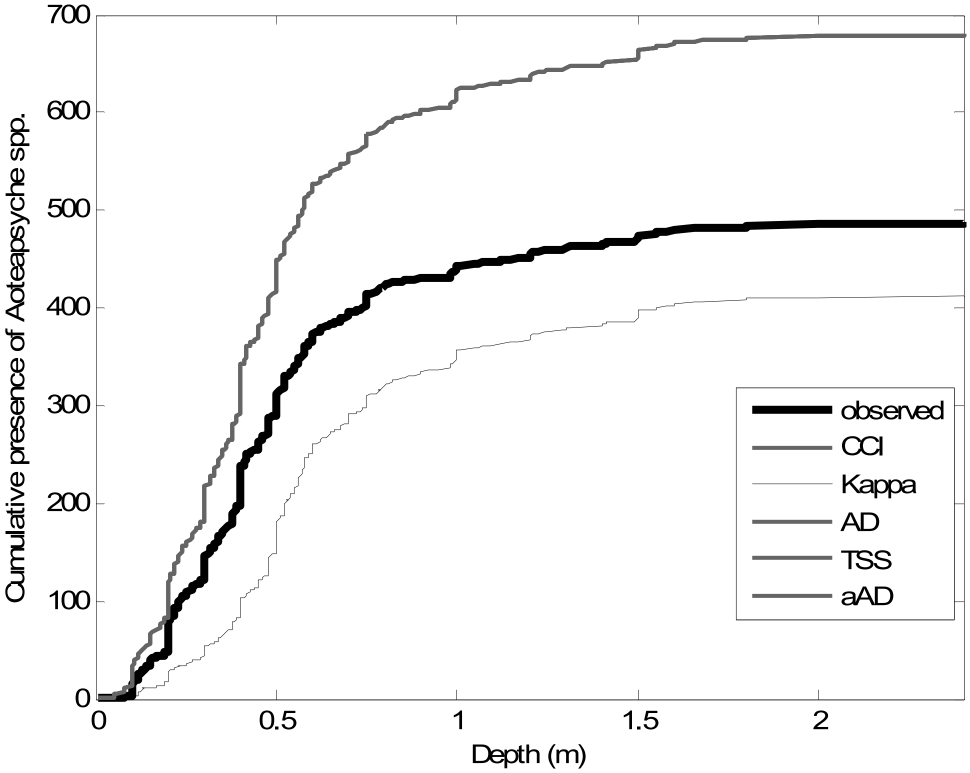


(c)

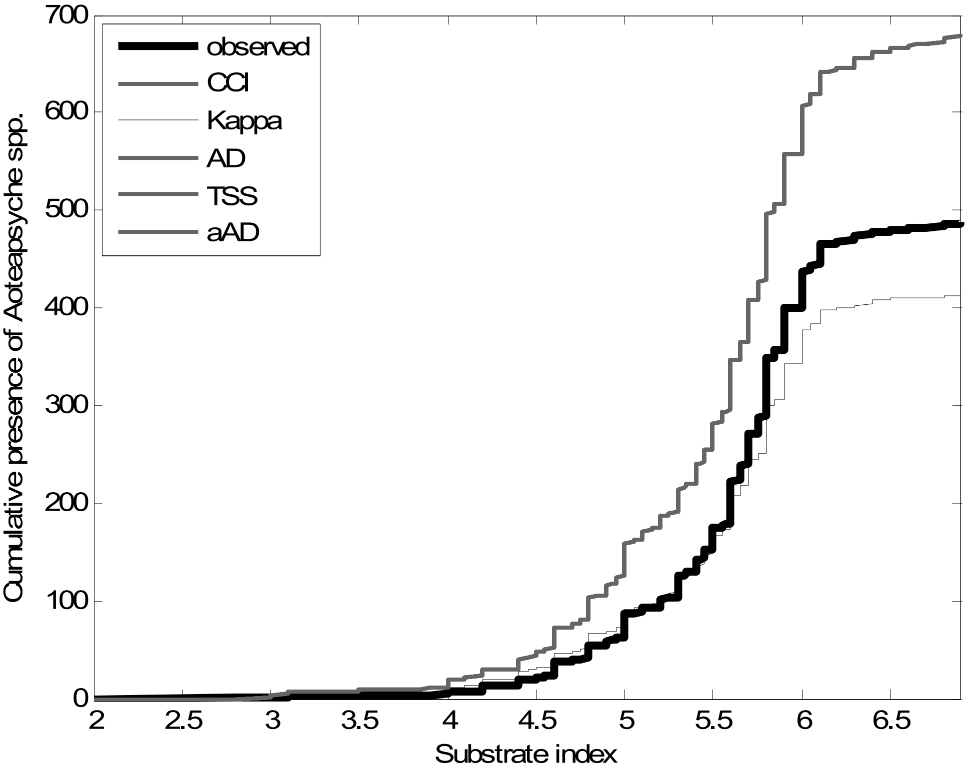
Fig. 9.3. Cumulative plots of the observations and of the rule base predictions of *Aoteapsyche spp.* obtained after training based on *aAD* with *α* varying between 0 and 1 in steps of 0.05. Predictions were averaged over the ten folds and cumulated according to their corresponding depth (a), flow velocity (b) or substrate index (c) values. Training was performed on a modified version of the original dataset with a prevalence of 0.20. Values of *α* resulting in the same core rule base, are indicated by the same line type.


observed
CCI
Kappa
AD
TSS
aAD
Cumulative presence of Aoteapsyche spp.
Flow velocity (m/s)
0
100
200
300
400
500
600
700
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6

(a)


observed
CCI
Kappa
AD
TSS
aAD
Cumulative presence of Aoteapsyche spp.
Depth (m)
0
100
200
300
400
500
600
700
0
0.5
1
1.5
2

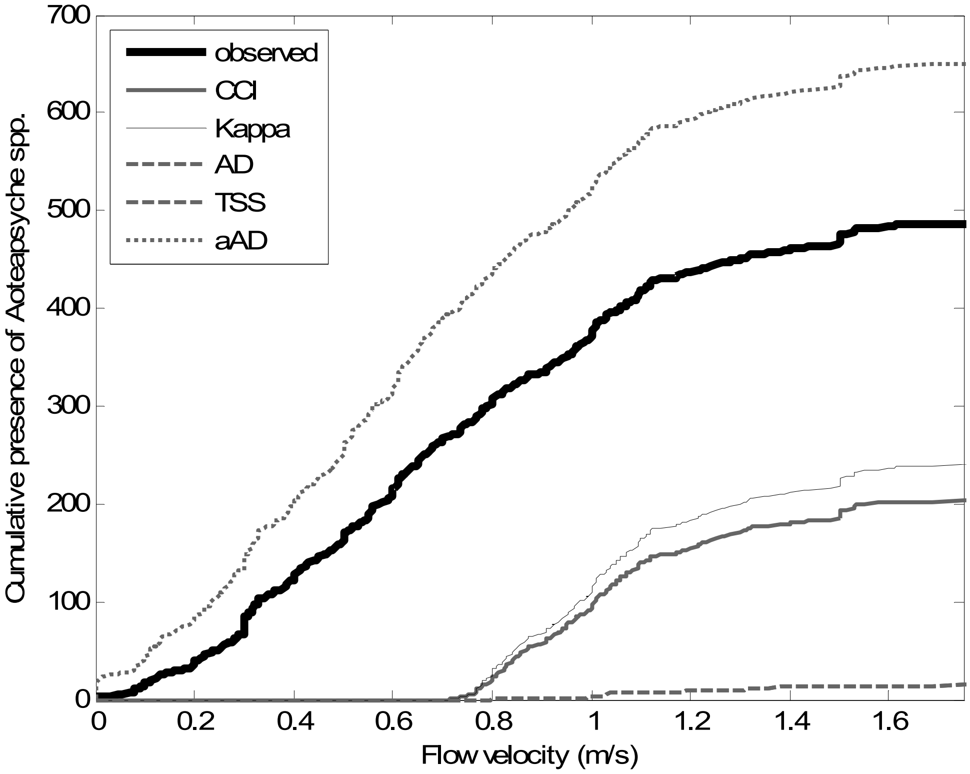
(b)



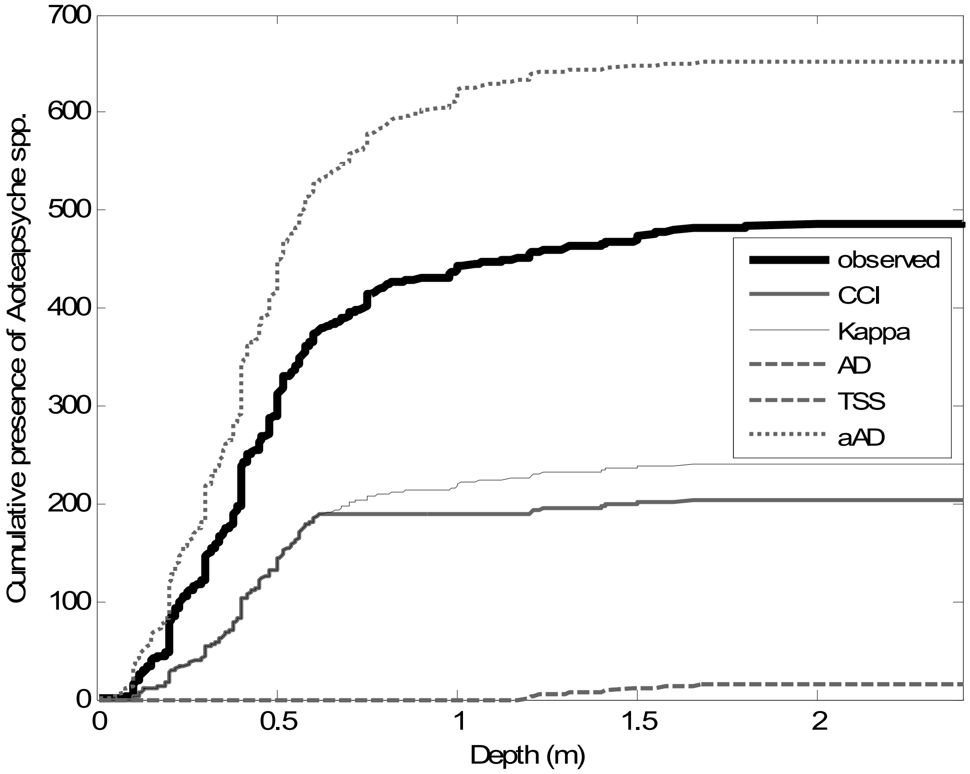
(c)

Fig. 9.4. The cumulative predictions of *Aoteapsyche* spp. for the variables flow velocity (a), depth (b) and substrate index (c) obtained after training based on the percentage of correctly classified instances (*CCI*), *Kappa*, the average deviation (*AD*), the true skill statistic (*TSS*), and the adjusted average deviation (*aAD*) with α = 0.65. Training was performed on the original dataset.

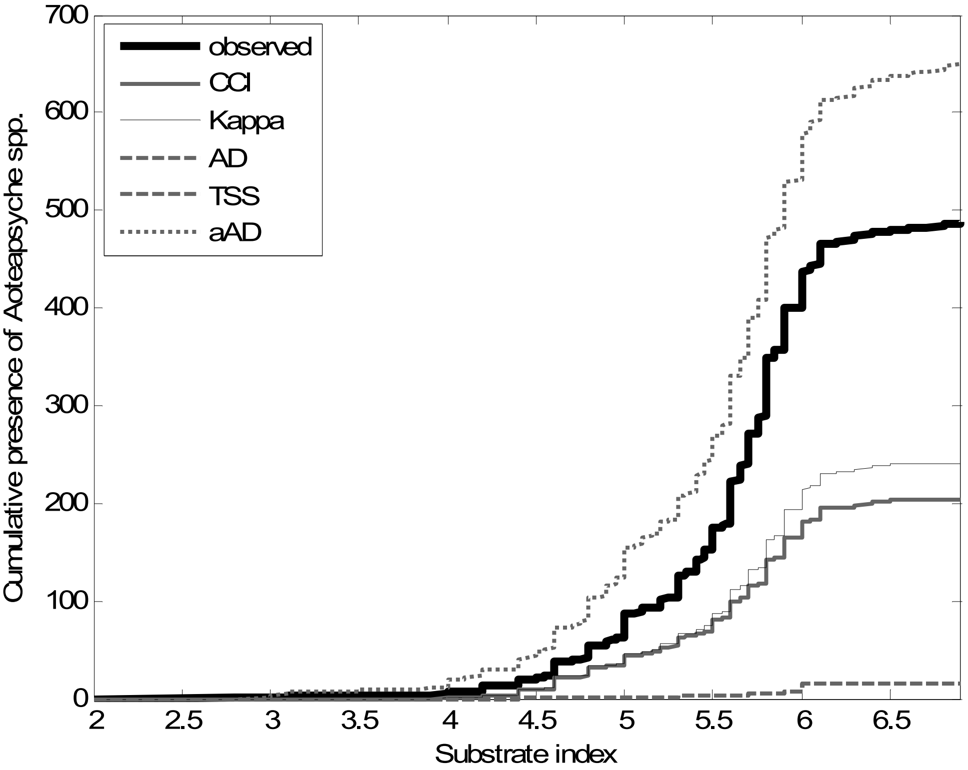
Comparison of the training based on *aAD* with training based on four other performance criteria revealed that at the original prevalence, all criteria but *Kappa* produced rule bases of which all consequents were ‘present’ (Fig. 9.4). Training based on *Kappa* underestimated the observations, especially in the 0.2 – 0.5 m depth range and the 0.1 – 0.7 m/s flow velocity range. Although false classifications appeared to occur over the whole substrate index range, similar results were found as those of the other input variables (Fig. 9.4). Comparison of the five performance criteria at a lower prevalence (0.20) of the training set, indicated that the *aAD* was the only criterion resulting in overestimating rule bases, whereas *AD* and *TSS* led to identical results at this lower prevalence (Fig. 9.5). Moreover, this similarity between *AD* and *TSS* is in line with the results obtained at a higher prevalence of the training set (Fig. 9.4). To calculate the *aAD* at both prevalences, the maximal value of α was chosen which resulted in the first rule base that overestimated the observations. Consequently, this maximal value of α was 0.65 at a prevalence of 0.71 and 0.2 at a prevalence of 0.2, which could also be observed in Fig. 9.1.


700
600
500
400
300
200
100
0
Cumulative presence of Aoteapsyche spp.
observed
CCI
Kappa
AD
TSS
aAD
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Flow velocity (m/s)

(a)


700
600
500
400
300
200
100
0
Cumulative presence of Aoteapsyche spp.
observed
CCI
Kappa
AD
TSS
aAD
0
0.5
1
1.5
2
Depth (m)

(b)

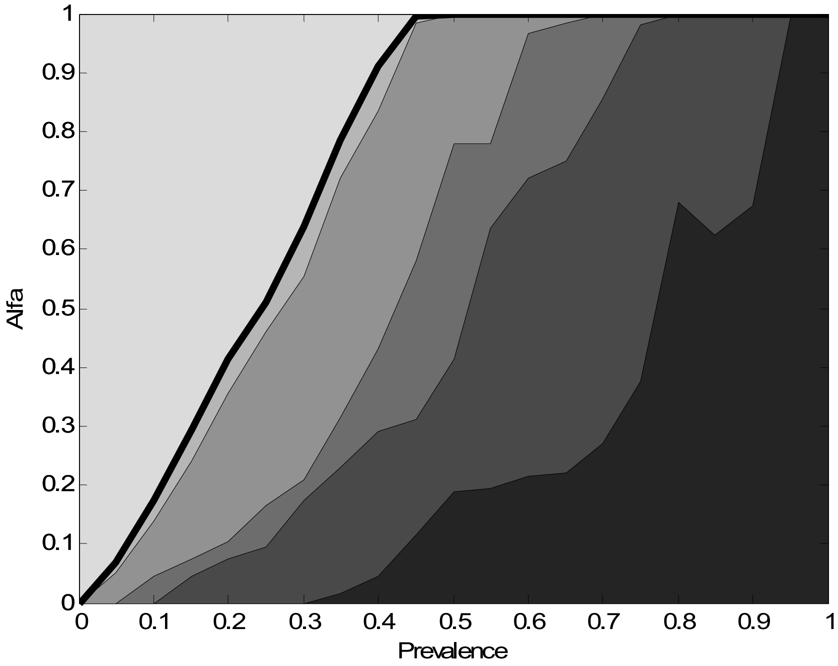


(c)

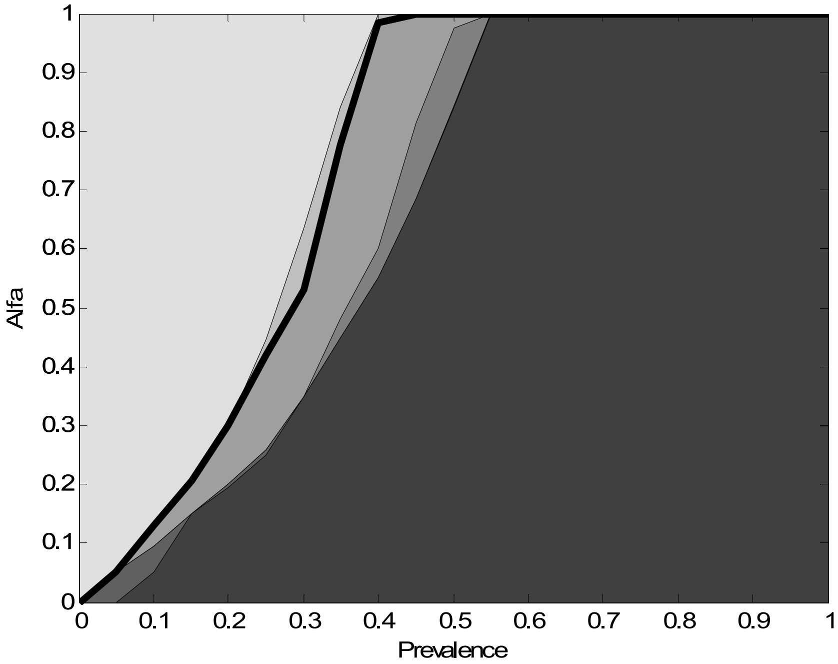
Fig. 9.5. The cumulative predictions of *Aoteapsyche* spp. for the variables flow velocity (a), depth (b) and substrate index (c) obtained after training based on the percentage of correctly classified instances (*CCI*), *Kappa*, the average deviation (*AD*), the true skill statistic (*TSS*), and the adjusted average deviation (*aAD*) with α = 0.20. Training was performed on a modified version of the original dataset with prevalence 0.20

9.3.2 Large brown trout and large rainbow trout

For large brown trout and rainbow trout, the number of model parameters was lower and the results were more consistent than for caddis fly. Consequently, fewer rule bases were obtained after training and no clustering was needed (Fig. 9.6). For both species, 6 different rule bases were found, although for large rainbow trout, some rule bases were only obtained within some limited α ranges. For large rainbow trout, all values of α which delimited the rule bases were increasing monotonously with the prevalence, whereas for brown trout some oscillations occurred (Fig. 9.6) at higher prevalence values. The number of 'present' consequents in the final rule base was negatively correlated with the value of the α parameter at each prevalence of both species (Fig. 9.6). For large brown trout, the second rule base (starting from the top left corner in Fig. 9.6a) first overestimated the observations, whereas for rainbow trout, three different rule bases first overestimated the observations, depending on the value of α (Fig. 9.6b).



(a)



(b)

Fig. 9.6. The 6 different rule bases obtained from rule base training based on *aAD* with varying α, averaged over the ten training folds at each prevalence of large brown trout (a) and large rainbow trout (b). Training was performed based on 21 training sets with a prevalence ranging between 0 and 1 in steps of 0.05, which were derived from the original dataset. Areas with the same shade represent the same core rule base, while the brightness of an area reflects the number of 'absent' consequents in the rule base. The light area (left top) contains the rule base with no 'present' consequents, whereas the darkest area (right down) represents the rule base with no 'absent' consequents. The left boundary of each area connects the maximum values of α at which the rule base of this area was obtained. The black line indicates the maximum values of α at which the first rule base was obtained that overestimated the observations.

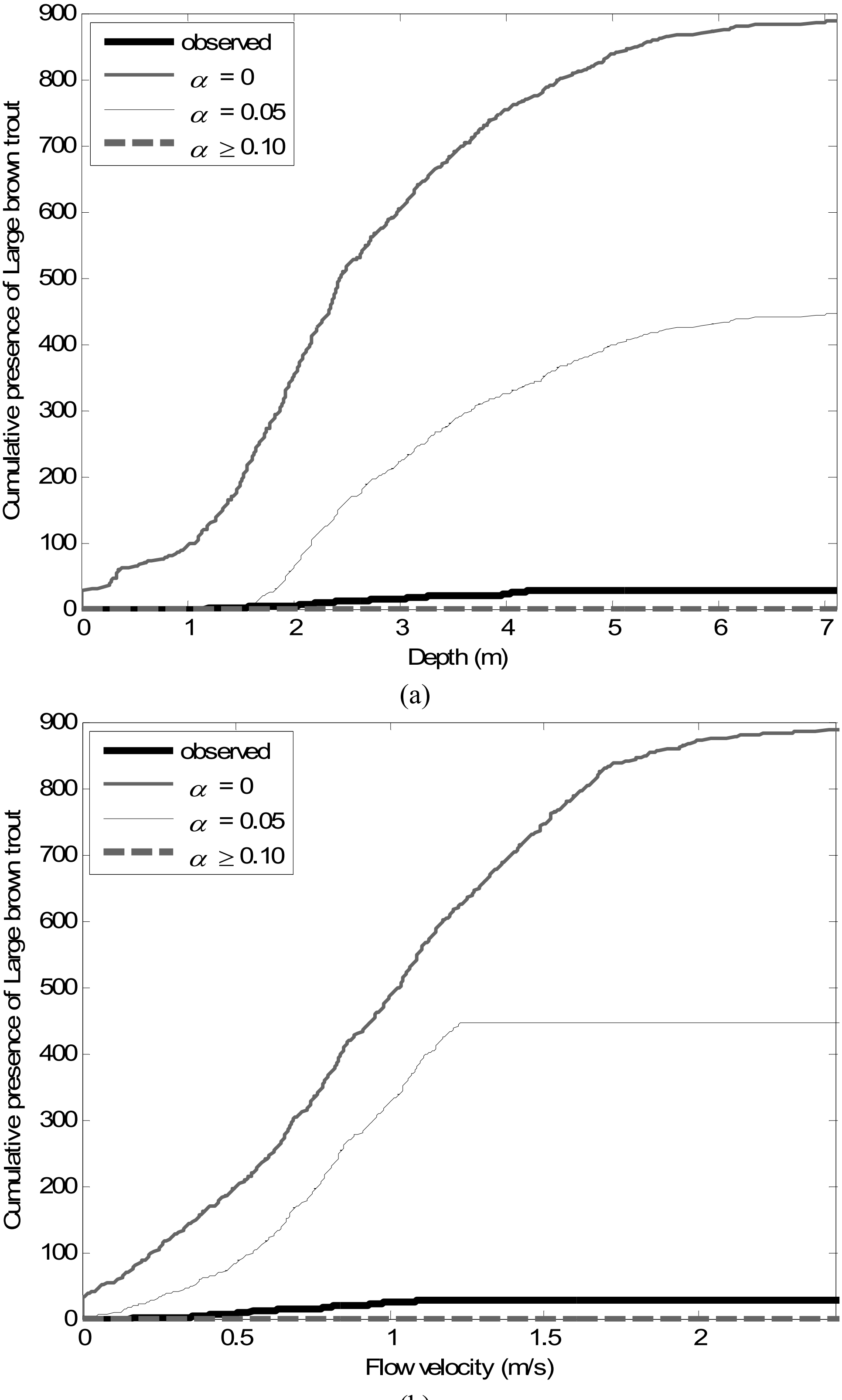
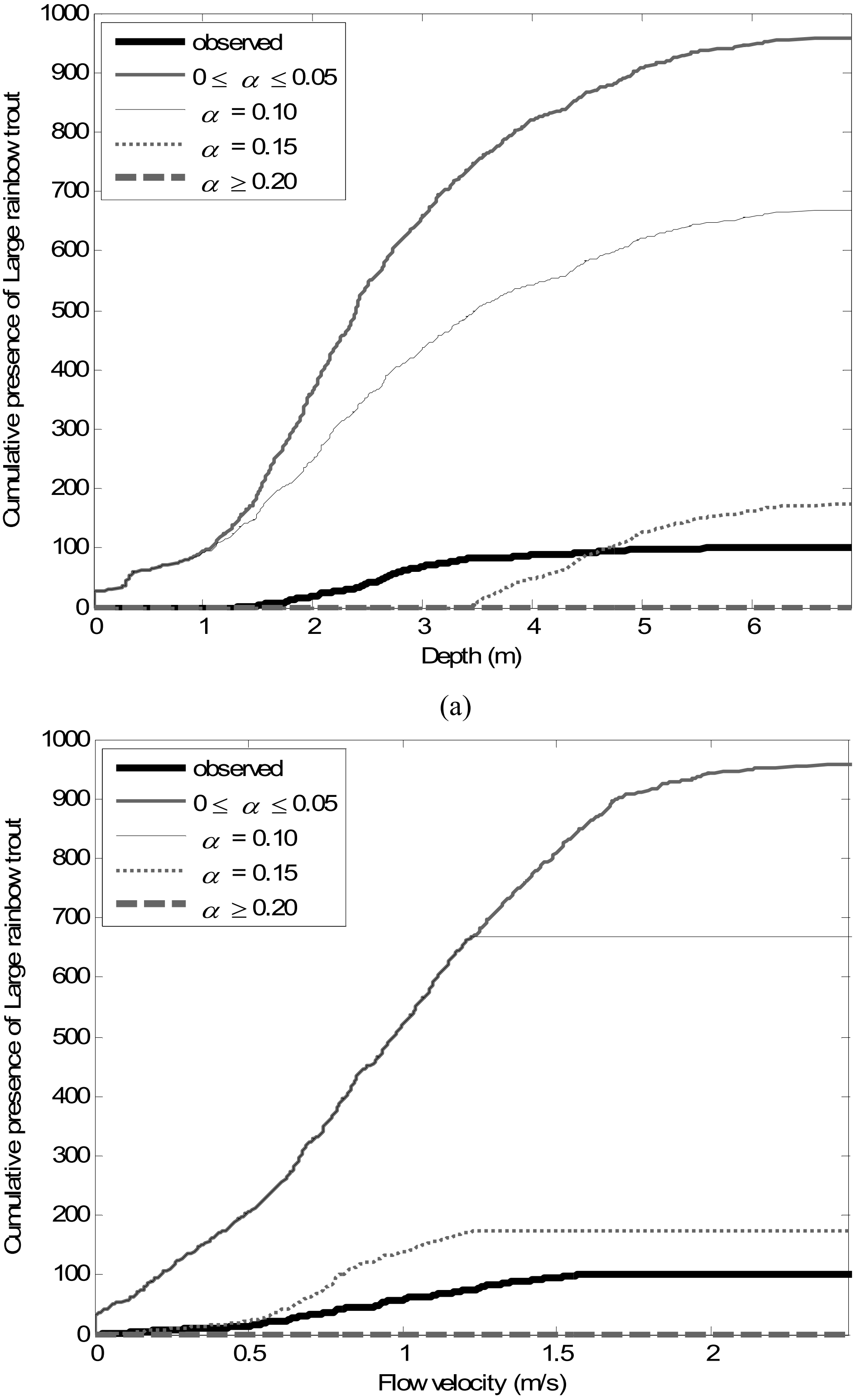


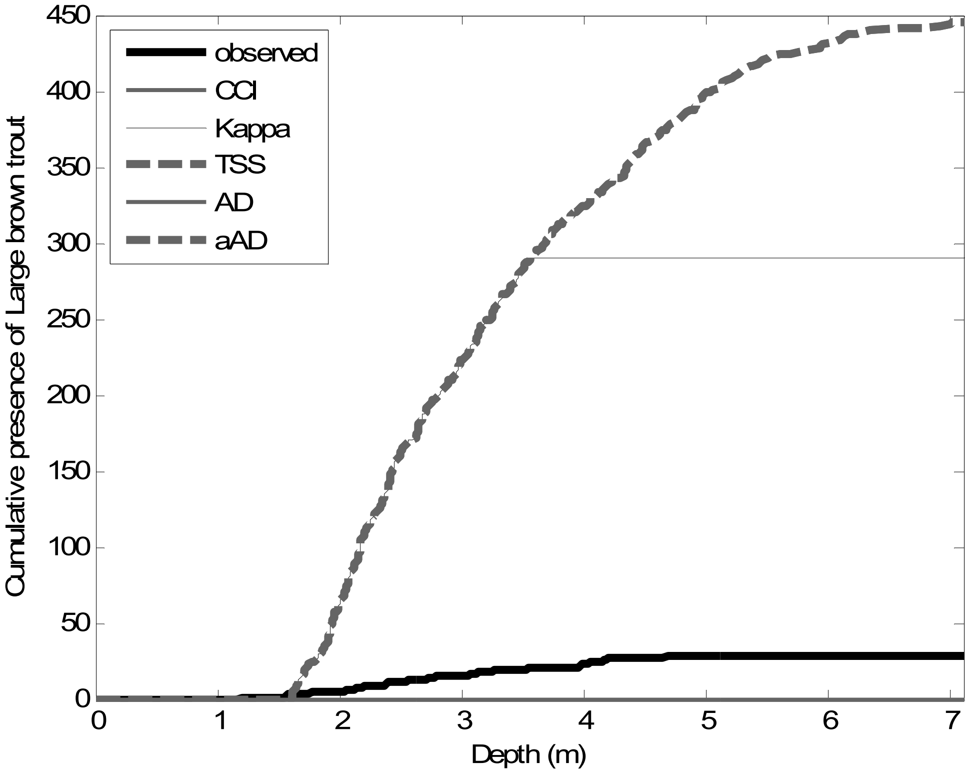
Fig. 9.7. Cumulative plots of the observations and of the rule base predictions of large brown trout obtained after training based on *aAD* with values of *α* varying between 0 and 1 in steps of 0.05. Predictions were averaged over the ten folds and cumulated according to their corresponding depth (a) or flow velocity (b) values. Training was performed on the original dataset with a prevalence of 0.03. Values of *α* resulting in the same core rule base, are indicated by the same line type.



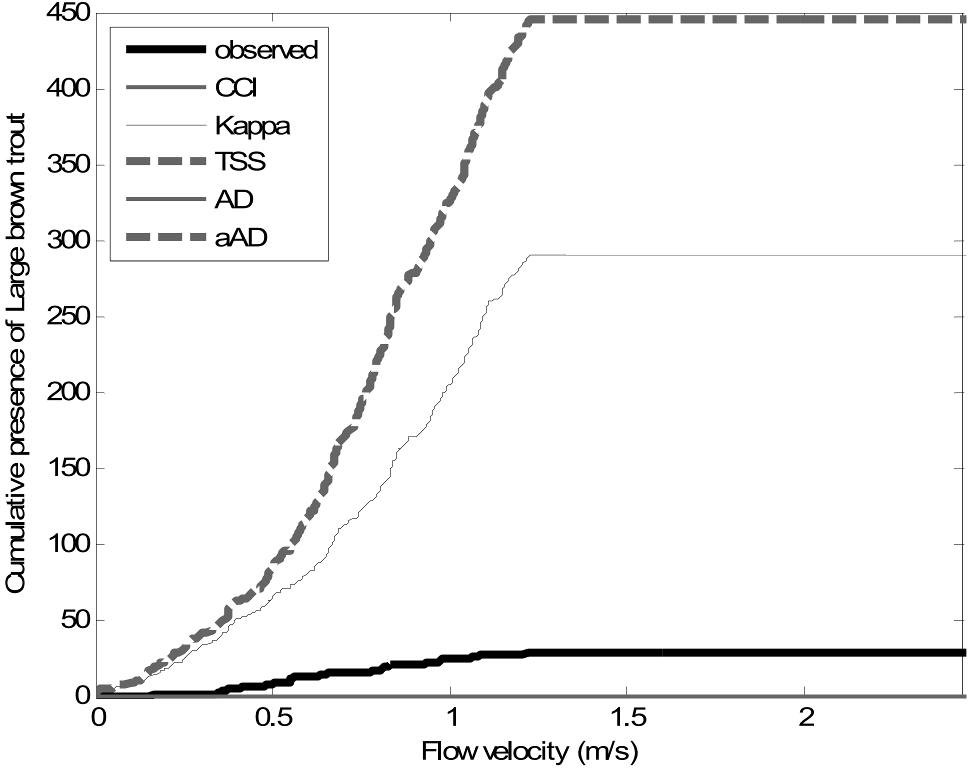
(a)

(b)

Fig. 9.8. Cumulative plots of the observations and of the rule base predictions of large rainbow trout obtained after training based on *aAD* with values of *α* varying between 0 and 1 in steps of 0.05. Predictions were averaged over the ten folds and cumulated according to their corresponding depth (a) or flow velocity (b) values. Training was performed on the original dataset with a prevalence of 0.11. Values of *α* resulting in the same core rule base, are indicated by the same line type.

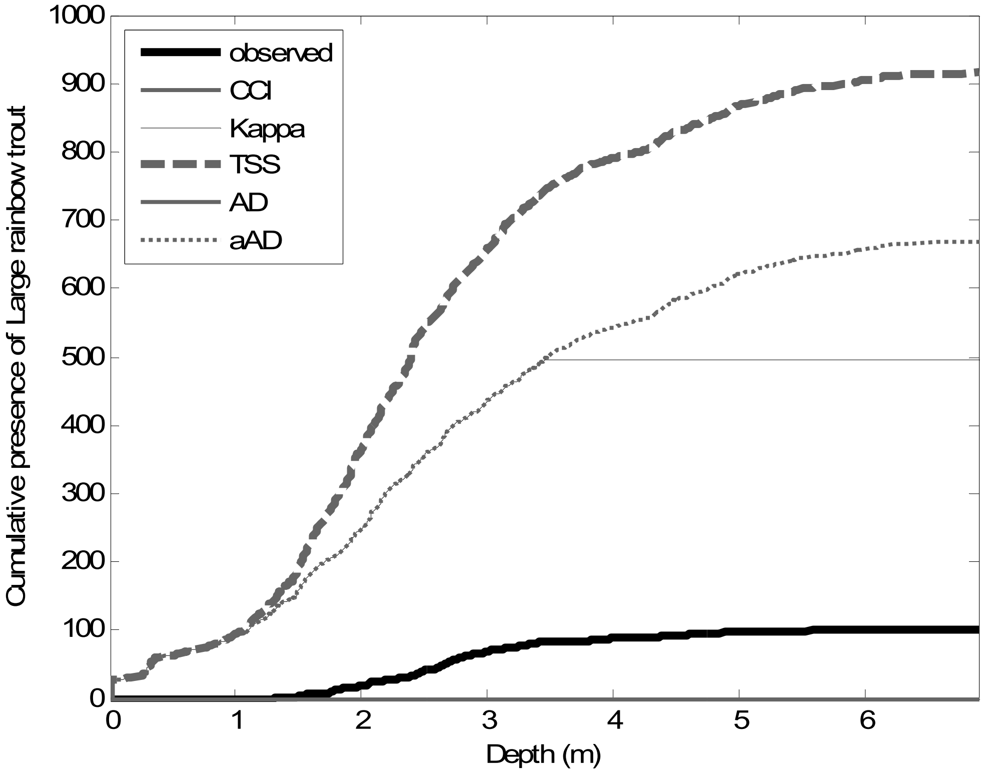


(a)

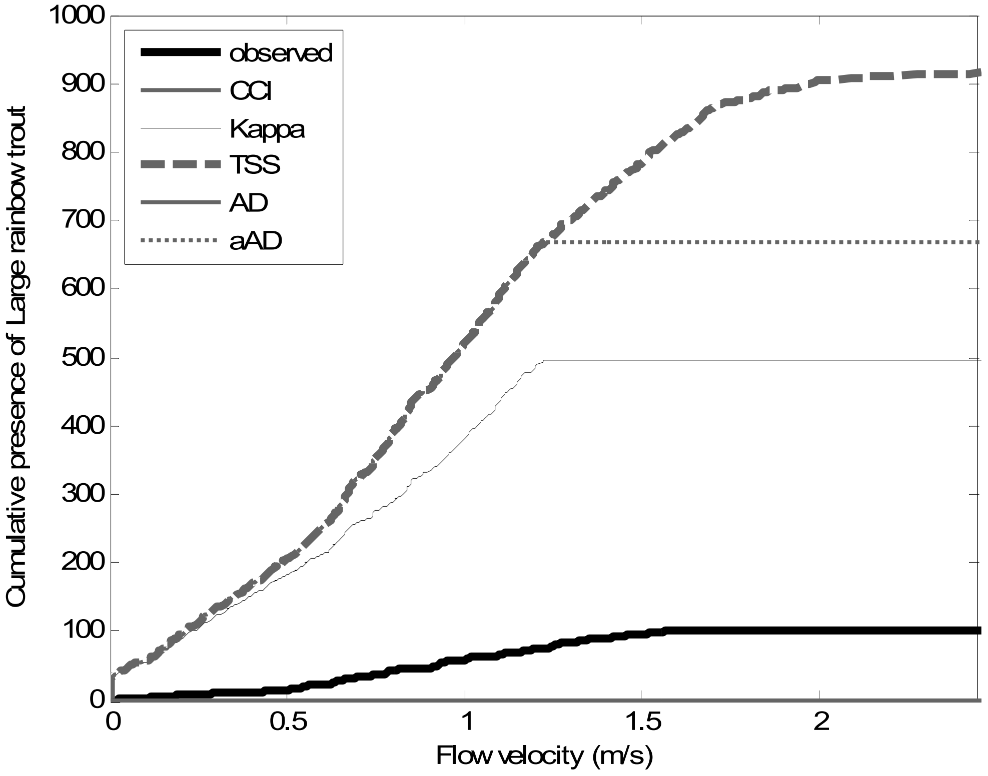


(b)

Fig. 9.9. The cumulative predictions of large brown trout for the variables depth (a) and flow velocity (b) obtained after training based on the percentage of correctly classified instances (*CCI*), *Kappa*, the average deviation (*AD*), the true skill statistic (*TSS*), and the adjusted average deviation (*aAD*) with α = 0.05. Training was performed on the original dataset.



(a)



(b)

Fig. 9.10. The cumulative predictions of large rainbow trout for the variables depth (a) and flow velocity (b) obtained after training based on the percentage of correctly classified instances (*CCI*), *Kappa*, the average deviation (*AD*), the true skill statistic (*TSS*), and the adjusted average deviation (*aAD*) with $\alpha = 0.15$. Training was performed on the original dataset.

The plots of the cumulative predictions of large brown trout indicated that the most accurate rule base might be found at values of *α* between 0.05 and 0.10 (Fig. 9.7), and between 0.15 and 0.20 for large rainbow trout (Fig. 9.8). For both species, model training based on the *AD* and *CCI* led to identical underestimating rule bases, while training based on *TSS*, *aAD* or *Kappa* resulted in rule bases that overestimated the observations (Figs. 9.9 and 9.10). For large brown trout, the same core rule bases are obtained after training based on *TSS* and on *aAD* (Fig. 9.9), whereas for rainbow trout, the rule base obtained after training based on *TSS* was overestimating the observations more than the one obtained after *aAD* training (Fig. 9.10). In contrast to the caddis fly results, training based on *Kappa* led to the most accurate rule bases for both species (Figs. 9.9 and 9.10).

9.3.3 Comparison of different values of *α*

The first values of *α* at which an overestimating rule base was obtained after model training were represented by a solid black line in Fig. 9.6. To analyse whether an universal relation between the value of *α* and the prevalence exists, Fig. 9.11 compares these values with values from a previous study on European grayling (*Thymallus thymallus* L.) in the Aare River in Switzerland. The three fish species show similar curves, whereas the value of *α* of caddis fly are significantly lower than those of the fish species.

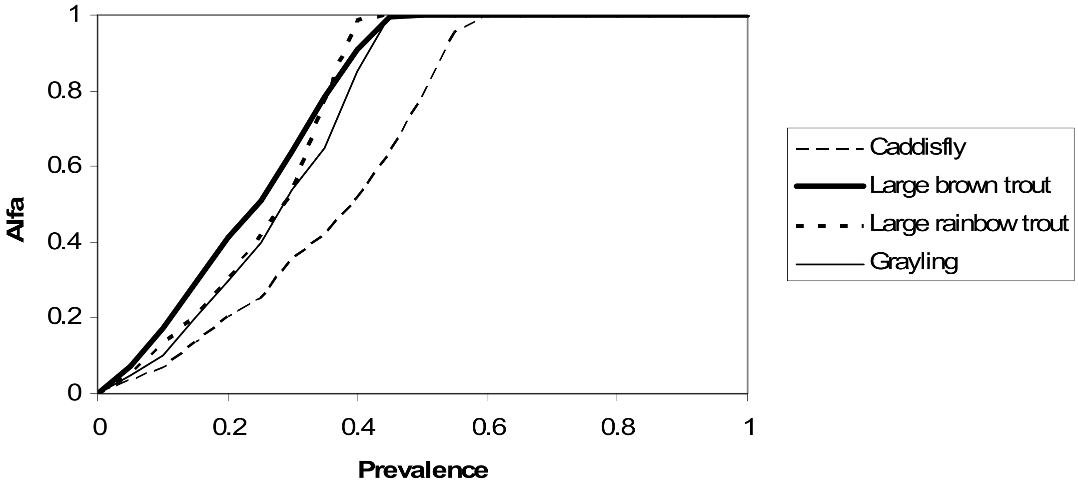


Fig. 9.11. The values of *α* at which the first rule base which overestimated the observations was obtained after model training for caddis fly, large brown trout, large rainbow trout and grayling. The results of grayling were derived from a previous study. All values were obtained by model training on the original datasets.

9.4 Discussion

The dependency between α and the training set prevalence also illustrates the effect of the training criterion on the final model. These results are in line with Chapters 6 to 8, which indicate that different training performance criteria could lead to different final models. An optimal parameter value or training performance criterion could be found by applying sensitivity analysis, but a more important problem with these flexible performance criteria could be the difficulty to say which models are better (Glas et al., 2003; Vaughan and Ormerod, 2005). The results in this chapter showed that comparison of the shapes of the cumulative prediction curves may provide an indication of the ecological relevance of the different optimised models.

The adjusted average deviation allows modellers to implement the balance between omission and commission errors in their models, while the α parameter indicates the extent to which a model is trained to overestimate or underestimate the observations. However, the caddis fly results show that the α parameter may be unable to prevent a trained model from overestimating the observations at high prevalence values of the training set. Therefore, further research should indicate if values of α higher than one may lead to underestimating models and thus improve the flexibility of the *aAD* training. Specifically, it could be shown theoretically that values of α higher than one stimulate underprediction. Another aspect that may affect model results is the α step size which is applied in the sensitivity analysis. The presented results show that there may be a substantial difference between two consecutive rule bases, especially between the last underestimating rule base and the first overestimating rule base. Previous research showed that applying smaller α step sizes may decrease this gap. However, this approach may only be appropriate if the number of model parameters allows sufficient fine tuning of model results. In this study, the number of model parameters was relatively low, which may have restricted the accuracy of the final model.

For all studied species, the results suggest that the optimal value of α for model training depends on the prevalence of the training set. However, this relation might depend on the prevalence of the original dataset from which the different training sets are derived. Specifically, this prevalence may reflect the ‘true’ model which is embedded in the original dataset. The prevalence of the caddis fly dataset is much higher than that of the two trout datasets, and thus the values of α needed to obtain an overestimating rule base may be lower for caddis fly than for trout. Moreover, the prevalence of the original dataset also affects the size of the training sets with adjusted prevalences which are derived from the original dataset. For the trout species, the training sets with a high prevalence were significantly smaller than those with a low prevalence, whereas for caddis fly, the training sets with higher prevalences were larger than those with a low prevalence. Previous studies indicated that the sample size of the training set may affect the results of the training procedure (Welsh, 1996). Consequently, further research should reveal if the size of the training set also affects the relation between the optimal α and the prevalence.

The strong selection of coarse substrata by *Aoteapsyche* concurs with earlier studies (Jowett and Richardson, 1990; Jowett et al., 1991) and is observed for many other benthic invertebrate species (Minshall, 1984). Although many invertebrates have an upper velocity tolerance limit, above which velocity exceeds the swimming or holding ability of the organism or mobilises the substrate on which the organism lives, individuals may be able to tolerate high mean water column velocities if friction and coarse substrata provide lower-velocity conditions near the bed. Despite the overestimation of the observations, the predictions of the models obtained after training based on *aAD* are similar to the observations of *Aoteapsyche* at the prevalence of the original dataset. Specifically, these models indicate that caddis fly approximately prefers depths between 0.2 and 0.5 m, flow velocities between 0.3 and 1.1 $m.s^{-1}$ and a substrate index between 5 and 6. These results are in line with previous results from a model study on the same datasets by Jowett and Davey (2007), who applied generalised additive models to predict the occurrence of the same species as those in this study. The model results obtained from training based on *aAD* and *Kappa* at lower prevalences (Fig. 9.5) show similar trends as those at the original prevalence (Fig. 9.4). However, training based on the other three criteria led to ecologically less relevant models at lower prevalences. For instance, training based on *AD* and *TSS* resulted in a model which predicted *Aoteapsyche* to occur only at depths higher than 1.2 m and at flow velocities higher than 1 m.s^{-1} (Fig. 9.5). Consequently, even though different training performance criteria may result in similar model predictions within the same prevalence ranges, they may lead to diverging model predictions at prevalence values outside these ranges (Fig. 9.5). Jowett and Davey (2007) also suggested that the relation between depth and *Aoteapsyche* occurrence may vary among different rivers, but this effect was not considered in this chapter since the number of samples available for each river was limited. Further research could reveal if this river effect is significant by applying the presented approach to larger datasets of different rivers.

For large brown trout, the *CCI* and *AD* model predicted brown trout to be absent in the studied stretch, whereas the *Kappa*, *TSS* and *aAD* model significantly overestimated the observations. However, *Kappa* predicted large brown trout to occur in a narrower depth range, whereas the *TSS* and *aAD* models predicted brown trout occurrence in a wider depth range than the observed range (Fig. 9.9). More similar results were obtained for flow velocity, where the *Kappa*, *TSS* and *aAD* models all predict brown trout to occur in the observed range, with the *TSS* and *aAD* models overestimating the observations more than the *Kappa* model. Jowett and Davey (2007) argue that the 29 large brown trout observed in the Clutha River were too few to produce a robust habitat suitability model but were probably sufficient to allow a qualitative comparison with existing habitat suitability curves. Although no habitat suitability criteria have been developed for brown trout in other large (> 100 m /s) New Zealand rivers, the results of this study reflected the range of preferred velocities (0.35 – 0.6 m.s^{-1}) reported in smaller New Zealand rivers (Hayes and Jowett, 1994). The preference for water deeper than 1.5 m was consistent with the habitat preferences reported by Hayes and Jowett (1994), who considered any depths greater than 0.5 m to be ideal brown trout habitat.

Similar to the large brown trout results, model training based on *CCI* and *AD* led to identical rule bases that predict large rainbow trout to be absent in the studied river. However, in contrast to brown trout, different results were obtained after training based on *TSS* and *aAD*. Both *Kappa* and *aAD* predicted rainbow trout to occur in a narrower flow velocity range than the observed range, while *Kappa* also predicted the preferred depth range to be narrower than the observed range. However, all three models reflected the preferred water velocity of large rainbow trout in the Clutha river (0.7 – 1.3 m.s^{-1}). This range was considerably higher than that reported in North American studies (Bovee, 1978; Leclerc, 1983; Baltz and Moyle, 1984; Raleigh et al., 1984; Suchanek et al., 1984; Hill and Hauser, 1985; Moyle and Baltz, 1985; Cochnauer and Elms-Cockrum, 1986; Lambert and Hanson, 1989; Thomas and Bovee, 1993; Pert and Erman, 1994). Most of these studies indicated optimal suitability at low velocities (< 0.4 m.s^{-1}), probably because the rivers (flows < 5 m .s^{-1}) and fish (typically 15 – 30 cm) were relatively small. Higher preferred velocities were reported in studies of adult rainbow trout in larger rivers such as the South Platte River, in which fish preferred flow velocities of 0.5 – 0.6 m.s^{-1} when actively drift feeding through a wide range of depths and velocities at flows of 7 – 17 m .s^{-1} (Thomas and Bovee, 1993). Similarly, adult rainbow trout in the Tongariro River (flow = 30 m .s^{-1}) preferred velocities of 0.5 – 0.7 m.s^{-1} (Jowett et al., 1996). In large rivers in Quebec, preferred velocities for rainbow trout ranged between 0.5 and 0.9 m.s^{-1} (Leclerc, 1983). The minimum preferred velocity in these three studies (0.5 m.s^{-1}) concurs with the observations in the Clutha river and with the results of the *TSS*, *aAD* and *Kappa* models. The maximum preferred velocity is more variable, however, and the upper limit in the Clutha river was the highest reported to date, as were the preferred depths for large rainbow trout in the Clutha river (Jowett and Davey, 2007). Preferences for depth in the aforementioned studies were more variable than for flow velocity, but generally were greater than 1 m (Leclerc, 1983; Thomas and Bovee, 1993; Jowett et al., 1996), which was also reflected in the results of the *TSS*, *Kappa* and *aAD* models.

Given the trout habitat preferences reported in this study, the previous section showed that similar criteria can be found in other rivers for trout of similar size and activity. In addition, the size and gradient of the river may also influence the habitat preferences by influencing the available habitat types. Specifically, if higher velocities and the associated energetic advantages were available, trout in low-velocity rivers may prefer these conditions (Jowett and Davey, 2007). Habitat preferences should thus be determined from observations of habitat use in a variety of habitats and rivers to guarantee the applicability of these preferences to a broad range of conditions. Jowett and Davey (2007) provide an example of Salmonids, which display three main feeding behaviours in rivers: cruise feeding, benthic browsing and drift feeding. They argue that piscivory and habitat requirements for these behaviours differ, which is illustrated by the predominant drift feeding in high- to moderate-gradient river reaches, whereas benthic browsing and cruise feeding occur in low-gradient, slow reaches. Consequently, one would ideally develop and use drift feeding, benthic browsing and resting habitat preferences. However, Jowett and Davey (2007) highlight that focus usually is on the habitat use known to have the highest flow requirements. They agree that in trout rivers, this will be adult drift-feeding habitat in combination with benthic invertebrate (food-producing) habitat (Jowett and Davey, 2007).

The presented results suggest that *Kappa* may be an appropriate performance criterion for robust model training, as was already assumed in Chapter 5. Model training based on *Kappa* leads to results which are relatively accurate, whereas the results of model training based on *TSS* may be more variable. Due to their strong dependency on the training set prevalence, the results in this chapter suggests that *AD* and *CCI* may be less appropriate for model training. Despite its user-friendliness, *Kappa* does not allow the model developers to focus on certain aspects of model performance and thus provides less control on the model training process than an adjustable performance criterion such as the *aAD*. This chapter may thus contribute to the transparency of the model training process and to an increasing awareness of the strengths and weaknesses of the final model. Not only could this improve the reliability of ecological models in general, but also enhance the application of these models in ecosystem management.

CHAPTER 10

General conclusions and further research

10.1 General conclusions

This section gives an overview of the main conclusions of the research concerning expert knowledge-based versus data-driven fuzzy models, the performance criteria applied to train and evaluate models, and the strengths and weaknesses of prevalence-adjusted model performance assessment.

10.1.1 Expert knowledge-based versus data-driven fuzzy models

Over the past decades, aquatic habitat suitability models have increasingly received attention due to their wide management applications in the context of biogeography, conservation biology and climate change studies. Ecological expert knowledge has been applied frequently in such models to link environmental conditions to the habitat suitability of aquatic species. Since the formalisation of problem-relevant human expert knowledge is often difficult and tedious, data-driven machine learning techniques may be helpful to derive expert knowledge from ecological datasets.

In Chapter 5, both fuzzy expert knowledge-based and data-driven habitat suitability models were developed and the performance of these models was compared. The results suggested that data-driven models may perform better than the expert knowledge-based models. This lack of consistency of ecological expert knowledge may hamper the application of expert models. Therefore, data-driven model development may complement the expert knowledge approach which is often used in habitat suitability modelling of aquatic species.

A habitat suitability modelling study of spawning European grayling in the Aare river (Switzerland) in Chapter 7 confirms these findings. The results showed a substantial gap between the data-driven and the expert knowledge-based predictions, while the first predictions approximated the observations more accurately. Moreover, the spawning habitats of grayling in the Aare river appeared to be deeper than those reported in previous studies. Possible explanations for the contrasting depth preferences may be differences in topology of the sampled rivers or ecological factors such as competition or predation. Specifically, the shallower spawning areas in the Aare rivers are being disturbed regularly by swans, which may have caused a shift in depth preferences. Chapter 7 thus provides an example in which ecological expert knowledge should be complemented by observed data to develop reliable ecological models.

The presented method was applied on a mayfly species in a specific river basin, but could be easily applied to any species, river system and at different spatial and temporal scales. The flexibility of the model allows modellers to select habitat variables and to define their classification in terms of fuzzy sets specific to the study site, the research objectives or the data availability. The approach in this dissertation allows for a quantitative description of expert knowledge, and thus enables comparison of the data-driven habitat suitability predictions with available expert knowledge. Consequently model users could choose the most appropriate model based on the quality of the available data and expert knowledge. Data-driven rules may be more reliable if a substantial number of high quality samples is available, whereas expert knowledge based rules may reliably predict habitat suitability in study areas which are underrepresented in the available dataset. To enable visualisation of the predicted habitat suitability, the final fuzzy rules could be implemented in tools such as the MesoCASiMiR module of the CASiMiR modelling system (Mouton et al., 2006; Mouton et al., 2007). As such, the presented method may provide a valuable tool for river managers and stakeholders to select different restoration options and to implement their management strategies.

10.1.2 Performance criteria

Although Chapter 5 showed that data-driven model development may complement the expert knowledge approach in species distribution modelling of aquatic species, these data-driven models should be developed with great care. Specifically, this dissertation emphasized the crucial role of the assessment of model performance during model training and the evaluation of the final model. Although numerous studies on species distribution modelling focus on the role of performance criteria for evaluation of the final model, few authors have addressed the effect of these criteria on model training. Moreover, a systematic overview was lacking because most analyses of performance criteria have been empirical and only focused on specific aspects of these criteria. Chapter 6 provides such an overview which showed that different performance criteria evaluate a model (or its resulting confusion matrix) differently, which may be explained by the dependency of these criteria on the prevalence of the validation set. Specifically, if models derived from different datasets are being compared, the prevalence of these datasets may affect the value of the performance measures and consequently manipulate the results of the comparison. Similar problems may arise when the performances of a model on a training set and a validation set with different prevalences are compared. Chapter 6 showed theoretically that these effects only occur if the data are inseparable by an n-dimensional hyperplane, with n the number of input variables. Given this inseparability, different performance criteria focus on different aspects of model performance during model training, such as sensitivity, specificity or predictive accuracy. These findings have important consequences for ecological modelling because ecological data are mostly inseparable due to data noise and the complexity of the studied system. Consequently, it should be very clear which aspect of the model performance is evaluated, and models should be evaluated consistently, that is, independent of, or taking into account, species prevalence.

Chapter 6 also highlighted the relative importance which the performance criteria give to omission and commission errors as a possible explanation for the differing evaluation scores among performance criteria for the same model. Theoretical analysis revealed that performance criteria may value a perfect model equally, but yet focus on very different aspects of model performance. To guarantee reliable quantification of the model performance, it is suggested that prevalence-independent measures should be preferred, and that at least some of these measures should allow modellers to distinguish between omission and commission errors for both model training and evaluation. The appropriate performance criteria for evaluation of the final model and for model training should be chosen carefully corresponding to the ecological objectives of the optimised model. The practical implications of this chapter are clear. It provides further insight in the evaluation of ecological presence/absence models and attempts to assist modellers in their choice of suitable performance criteria.

The results of the theoretical analysis were confirmed in a first experimental analysis in Chapter 7. A heuristic nearest ascent hill-climbing algorithm for rule base optimisation was applied to construct a fuzzy rule-based habitat suitability model for spawning European grayling (*Thymallus thymallus* L.) in the Aare river (Bern, Switzerland). Optimisation of the fuzzy rule-based model was based on two different training criteria, the weighted Correctly Classified Instances (CCI_w) and Cohen's *Kappa*. The ecological relevance of the results was assessed by comparing the optimised rule bases with a rule base derived from ecological expert knowledge. The results showed that model training based on CCI_w and based on *Kappa* lead to different final models and thus confirmed the theoretical analysis described in Chapter 6. Specifically, in Chapter 6 it could be shown that these different final models could be attributed to the inseparability of the data. If the data would have been separable, all optimisation criteria would have led to the same results. Optimisation based on *Kappa* appeared to generate acceptable results and was easier to use than optimisation based on CCI_w because the latter required fine tuning of a weight parameter which accounted for the species prevalence. However, this parameter provided more transparency in the model training process by quantifying the degree to which overprediction had to be stimulated to result in accurate models. Specifically, this approach allowed modellers to adjust the model training process to the model objectives and to gain insight into this process due to the quantification of the adjustment.

10.1.3 Prevalence-adjusted model performance quantification

Although the CCI_w applied in Chapter 7 provided a first step towards prevalence-adjusted model performance quantification, this performance criterion cannot deal with the fuzzy output of a fuzzy classifier because it is based on the confusion matrix. Such performance criteria, for instance, are not sensitive to the position of the fuzzy sets where the wrong classification occurs, and valuable information may thus be lost by transferring this fuzzy output to the crisp output which is needed to generate the confusion matrix. Therefore, in Chapter 8 the adjusted average deviation (*aAD*) is introduced. This performance measure includes a parameter α which ranges between 0 and 1 and allows stimulation of overprediction or underprediction, depending on the prevalence of the training data set.

To analyse the strengths and weaknesses of this approach, it was applied on model training and the results were compared to those obtained after model training based on some of the most frequently applied performance criteria: *Kappa*, *CCI*, average deviation (*AD*) and true skill statistic (*TSS*). The results suggest that the *CCI*, *AD* and *TSS* criteria may be less appropriate for model training at extreme prevalence values. However, *TSS* and *Kappa* showed to be valuable criteria for the evaluation of the performance of the final model. Both accuracy-based measures, *CCI* and the *AD*, appeared to be less suitable for model evaluation because their values were correlated with the value of α. In Chapter 8 the results of Chapter 7 are confirmed by showing that the performance criteria which are applied for model training may significantly affect the results of the training process.

Finally, in Chapter 9 the relation was analysed between the parameter α and the training set prevalence for three other species in different New Zealand river systems: a caddis fly, large brown trout and rainbow trout. Like in Chapter 8, the *aAD* was implemented in a hill-climbing algorithm to optimise a fuzzy species distribution model for each species. For all studied species, the results proved that the optimal α value for model training depends on the prevalence of the training set. The results in Chapter 9 also confirmed that *AD* and *CCI* may be less appropriate for model training due to their strong dependency on the training set prevalence. Moreover, these results agreed that *Kappa* may be an appropriate performance criterion for robust model training. Model training based on *Kappa* led to results which were relatively accurate, whereas the results of model training based on *TSS* were more variable. Despite its user-friendliness, however, *Kappa* does not allow the model developers to focus on certain aspects of model performance and thus provides less control on the model training process than an adjustable performance criterion such as the *aAD*. Finally, Chapter 9 suggests that model training based on accuracy-based performance criteria may produce unrealistic models at extreme prevalences of the training dataset, whereas the *aAD* allows identification of more accurate and well-founded models. Moreover, comparison of the relation between the α parameter and the prevalence of the training data set for different species revealed this relation might be consistent for similar data sets and may depend on data or model characteristics.

An optimal parameter value or training performance criterion could be found by applying sensitivity analysis, but a more important problem with these flexible performance criteria could be the difficulty to say which models are better. Although applying a training parameter which minimises the difference between over- and underprediction may be a suitable rule of thumb, there is no straightforward answer to this question, because the quality of a model depends on its purpose and the context in which the model will be applied. Conservationists should thus distinguish between models which reliably or less reliably predict species distribution. Given inseparable ecological data, each performance criterion will focus on different aspects of model performance, and these aspects are reflected in the final model which is obtained after training based on this performance criterion. Moreover, in Chapter 6 it was shown theoretically that the aspects on which these criteria focus may vary with the prevalence of the training dataset. Consequently, conservationists and model developers should be aware of the different focus of the training performance criteria, and thus of the final model.

This dissertation may thus contribute to the transparency of the model training process and to an increasing awareness of the strengths and weaknesses of the final model. Specifically, the adjustable parameter in this criterion enables modellers to situate the optimised models in the heuristic search space and thus provides an indication of the ecological model relevance. Consequently, it may support modellers and river managers in the decision making process by improving model reliability. The habitat suitability models optimised in this dissertation are able to predict the effect of different impacts on the river system and to select the optimal restoration option. These models could thus be a valuable decision support tool for river managers and ease the discussion between stakeholders. As such, they may be a small step on the road towards more reliable species distribution models. Due to the universality and the flexibility of the approach, it could be applied to any other ecosystem or species, and may therefore be valuable to ecological modelling and ecosystem management in general.

10.2 Indications for further research

Fuzzy species distribution modelling

The input variables of the habitat suitability models in this dissertation were defined by fuzzy sets. This approach allowed the implementation of the ecological gradient theory into the models. Specifically, the transition between two consecutive variable classes was gradual and reflected the transition between different environmental conditions. This approach has particular advantages if the model is applied in new situations as was described in this dissertation. Although the membership functions of these fuzzy sets are often defined based on expert judgement, this dissertation described a fuzzy set optimisation procedure to ensure a uniform distribution of the input variables values over the fuzzy sets. However, these membership functions may influence the final model predictions. Therefore, further research could reveal the degree to which the model predictions are sensitive to the membership functions and thus enhance model robustness. Moreover, field validation or the creation of artificial species with data noise addition could indicate which fuzzy set approach leads to the most reliable habitat suitability model.

Role of expert knowledge in species distribution modelling

Although the implementation of expert knowledge in habitat suitability models may be useful, this dissertation showed that ecological data may significantly improve the reliability of the models. As mentioned in Chapter 2, these findings may be explained by the characteristics of the variables which affect habitat suitability. Depending on the situation, these variables may be an indirect or a direct gradient and thus the species response to these variables may change. Moreover, recent studies in aquatic habitat suitability modelling focus on the appropriate spatial and temporal scale level at which a model should be developed. Further research may also reveal whether ecological expert knowledge is consistent across these various scale levels. Although expert knowledge has often been referred to as static, the results in this dissertation demonstrate the dynamical character of expert knowledge. Therefore, this knowledge should be treated flexibly and should be adjusted for different species life stages, seasons, sampling sites, ... As such, temporal, population dynamical and spatial aspects may be incorporated in species distribution models and thus lead to more reliable models. A possible limitation of this approach is the reduction in model transparency due to the increase in model elements such as different modules for different seasons for example. However, the methods presented in this dissertation provide an appropriate framework for elaboration of knowledge which can easily be exchanged between experts. First, an initial rule base could be created based on the available data and on the fuzzy methodology in this dissertation. This rule base can then be inspected, used and modified by human experts applying expert knowledge or new insights to cover a wider range of environmental conditions.

Performance criteria in species distribution modelling

The forelaying dissertation is only a small step in the search towards prevalence-adjusted performance quantification and suitable performance criteria. Depending on the modelling purpose, new performance criteria could be created or appropriate performance criteria could be selected. The presented performance measure selection for model training based on the cumulative predicted presence of this model might also be applied to training of other data-driven models such as artificial neural networks, decision trees and generalised linear models. This could reveal whether similar relations between performance criterion values and training set prevalence are found for both fuzzy and other modelling techniques. Moreover, different performance measures should be compared to evaluate model performance after training. In line with the performance measure selection for model training, the optimal performance measure for model evaluation should be chosen depending on the classification problem and on the model purpose. The results in this dissertation also reinforce the view that categorisation of performance criterion values into classes from fair to almost perfect is arbitrary. Hence, the value of such criteria should be interpreted by comparison with other criteria values derived from models classifying the same data set or data sets with similar characteristics.

Relation between performance criteria and data characteristics In this dissertation, it was demonstrated that prevalence-adjusted performance quantification may be an appropriate technique to develop reliable habitat suitability models. However, fine tuning of the parameter in the prevalence-adjusted performance criterion is needed. Although this adjustment could be performed easily by sensitivity analysis, the results of this dissertation suggest that there may be a general relation between the value of the α parameter in the aAD and the prevalence of the training data set. However, further research should indicate whether this relation also holds for other species in other ecosystems. Moreover, the relation between the parameter and other data or model characteristics should be analysed to gain insight into the model development process. Finally, future research should reveal which performance measure is the most appropriate for model training, depending on the classification problem and on the model purpose.

Variable selection in fuzzy species distribution modelling
The approach in this dissertation allows selection of an appropriate performance criterion for model training and this criterion reliably quantifies the performance of the fuzzy species distribution model. Based on this performance, the importance of the different input variables in the model could be calculated with the procedure described in Chapter 4. Although this procedure was designed to generate a consistent ranking of the input variables, several factors may influence the variable selection results, such as the membership functions of the fuzzy sets, data characteristics or the modelling technique. Further research could indicate whether these factors affect the variable selection significantly and lead to further fine-tuning of the procedure.

10.3 Main contributions of this dissertation

The aim of this dissertation was to develop reliable species distribution models for aquatic species. Fuzzy modelling has proved to be an appropriate technique for incorporation of ecological aspects into these models. Therefore, several procedures were developed to improve the applicability of fuzzy models on species distribution modelling. As such, this dissertation contributed to both the species distribution modelling and the fuzzy modelling domain. These contributions are the following:

- the strengths and weaknesses of expert knowledge in species distribution modelling were illustrated and compared with those of a data-driven approach;

- a computational method to select variables in a fuzzy species distribution model was developed and applied;

- a fuzzy set optimisation method was developed and applied to improve the uniformity of the distribution of the input variable values over the fuzzy sets;

- a data-driven rule base optimisation procedure was introduced and applied for species distribution modelling on two benthic invertebrate and three fish species;

- an overview was provided of the different assets and pitfalls of the most frequently applied performance criteria in species distribution modelling. This overview was illustrated with both theoretical and empirical analyses;

- a new performance criterion was presented taking into account the fuzzy characteristics of the models and the difference between omission and commission errors.

Appendix

APPENDIX A

Analysis of *AD* and *aAD*

Chapter 6 provided a theoretical analysis of the effect of the training set prevalence on the final model for model training based on the most frequently applied performance criteria. In this appendix, this analysis is also performed for the average deviation *AD* and the adjusted average deviation *aAD*.

The change in *AD* and *aAD*, respectively Δ*AD* and Δ*aAD*, resulting from a change Δ*a* as described in the confusion matrix in Table 6.6, can be calculated based on the *AD* and *aAD* which are transformed to confusion matrix-based performance measures as described in Chapter 4. Since in this situation, *AD* is the inverse of *CCI*, the change in *AD*, Δ*AD*, resulting from a change Δ*a*, can be calculated as

$$\begin{aligned}\Delta AD &= AD^* - AD \\ &= \frac{(c - \Delta a + b + \Delta d)}{N} - \frac{(c+b)}{N} \\ &= \frac{(\Delta d - \Delta a)}{N}\end{aligned} \tag{A.1}$$

If Δ*AD* is smaller than zero, the adjustment of the model parameters will result in a better model and the optimisation algorithm will continue with this adjusted model. Consequently, the relation between Δ*AD* and prevalence is exactly the same as the relation between Δ*CCI* and prevalence.

The change in *aAD*, Δa*AD*, resulting from a change Δ*a*, can be calculated as

$$\begin{aligned}\Delta aAD &= aAD^{*} - aAD \\ &= \frac{(c - \Delta a + \alpha \cdot (b + \Delta d))}{N} - \frac{(c + \alpha \cdot b)}{N} \\ &= \frac{(\alpha \cdot \Delta d - \Delta a)}{N}\end{aligned} \tag{A.2}$$

If Δa*AD* is smaller than zero, the adjustment of the model parameters will result in a better model and the optimisation algorithm will continue with this adjusted model. In this situation,

$$\begin{aligned}\Delta aAD < 0 &\Leftrightarrow \frac{(\alpha \cdot \Delta d - \Delta a)}{N} < 0 \\ &\Leftrightarrow \alpha < \frac{\Delta a}{\Delta d}\end{aligned} \tag{A.3}$$

At high prevalences ($P \approx 1$), the likelihood that Δ*a* exceeds Δ*d* is relatively high, and thus the upper limit of *α* will be equal to or greater than one. If the prevalence is decreasing towards zero, the change that Δ*a* exceeds Δ*d* is also decreasing, which results in a decreasing upper limit of *α*. Specifically, the lower the prevalence, the lower *α* has to be to ensure that a change Δ*a* will lead to a better model.

Table A.1 shows the values of 9 different performance criteria for the 4 scenarios: the percentage of correctly classified instances (*CCI*; Fielding and Bell, 1997), Cohen's *Kappa* (Cohen, 1960), the true skill statistic (*TSS*; McBride and Ebert, 2000; Saseendran et al., 2002; Elmore et al., 2003; Allouche et al., 2006), the *NMI*, the specificity (*Sp*), the sensitivity (*Sn*). Based on the theoretical analysis described in this appendix, Table 6.4 could be extended by the Average Deviation (*AD*; Van Broekhoven et al., 2007) and the adapted Average Deviation (*aAD*) as follows:

Table A.1. The values of the 9 different performance criteria for the 4 scenarios, assuming that ln(0)=0. To calculate the odds ratio, a continuity correction was performed by adding 0.5 to each of the cells in the confusion matrix (Forbes, 1995; Vaughan and Ormerod, 2005). These assumptions have no effect on the characteristics of the presented criteria.

Criterion Scenario	*CCI*	*Sn*	*Sp*	*NMI*	*Kappa*	*TSS*	Odds ratio	*AD*	*aAD*
1	1	1	1	0	1	1	25	0	0
2	0	0	0	0	-1	-1	0	1	1+0.5α
3	0.75	1	0.5	0.23	0.5	0.5	5	0.25	0.25α
4	0.75	0.5	1	0.23	0.5	0.5	5	0.25	0.25

The performance criteria *CCI*, *NMI*, *Kappa*, *TSS*, odds ratio and *AD* do not distinguish between scenarios 3 and 4, whereas *Sn* and *Sp* allow differentiation between these two scenarios. However, *Sn* and *Sp* do not distinguish between scenario 1, and scenarios 3 and 4, respectively. The *aAD* is the only criterion which distinguishes between all four scenarios, given an appropriate value of the *α* parameter. Because the presented example is based on dichotomous presence-absence model scores, no threshold was applied to generate the confusion matrices. In this situation, *AD* equals 1-*CCI*. Consequently, Table 6.9 could be extended as follows:

Table A.2. Characteristics of the most frequently applied performance criteria for model training and evaluation. *NMI* = Normalised Mutual Information Statistic; *TSS* = True Skill Statistic; AD_{discr} = Average Deviation calculated from dichotomised predictions; *AUC* = Area Under the Curve; *CCI* = Correctly Classified Instances; *Sn* = Sensitivity; *Sp* = Specificity; AD_{cont} = Average Deviation calculated from continuous predictions; x = the characteristic fully applies to the performance criterion; - = the characteristic does not apply to the performance criterion; ? = the characteristic may apply to the performance criterion.

Performance criterion Characteristic	*Kappa*	*NMI*	Odds ratio	*TSS*	AD_{discr}	*AUC*	*CCI*	*Sn/Sp*	AD_{cont}
o quantifies the extent to which models correctly predict occurrence better than chance expectation	x	x	x	x	-	x	-	-	x
o depends on prevalence	x	?	?	x	x	?	x	x	?
o takes into account the complete information included in the confusion matrix	x	x	x	x	-	n.a.	-	-	n.a.
o does distinguish between omission and commission errors	-	-	-	-	-	-	-	-	-
o compensates for extreme prevalence values when applied on model training	x	-	-	-	-	-	-	-	-
o requires discretisation of model predictions by applying threshold values	x	x	x	x	x	-	x	x	-
o allows zero values in the confusion matrix	x	-	-*	x	x	n.a.	x	x**	n.a.
o is proportional (the same performance is found if all elements of the confusion matrix are divided by the same constant)	x	-	-	x	x	n.a.	x	x	n.a.
o is suitable for model training	x	-	x	x	x	-	x	x	x

* cannot be applied directly when both the number of false-positive predictions and false-negative predictions is zero; adding a constant value to each element of the confusion matrix changes the relative value of the odds ratio

** cannot be applied if the prevalence of the evaluation or training set is 0.

REFERENCES

Acreman, M. and Dunbar, M.J. (2004). Defining environmental river flow requirements - a review. *Hydrology and Earth System Sciences*, 8, 861-876.

Adriaenssens, V. (2004). Knowledge-based macroinvertebrate habitat suitability models for use in ecological river management. Ph.D. Thesis, Ghent University, Gent, Belgium, 296 pp. + annexes.

Adriaenssens, V., De Baets, B., Goethals, P.L.M. and De Pauw, N. (2004). Fuzzy rule-based models for decision support in ecosystem management. *Science of the Total Environment*, 319, 1-12.

Adriaenssens, V., Goethals, P.L.M. and De Pauw, N. (2006). Fuzzy knowledge-based models for prediction of *Asellus* and *Gammarus* in watercourses in Flanders (Belgium). *Ecological Modelling*, 195, 3-10.

Akaike, H. (1974). A New Look at Statistical-Model Identification. *IEEE Transactions on Automatic Control*, 19, 716-723.

Alfredsen, K. and Killingtveit, A. (1996). The Habitat Modelling Framework. A tool for creating habitat analysis programs. In: Leclerc, M., Capra, H., Valentin, S., Boudreault, A. and Cote, Y. (Eds.), Proceedings of the 2nd International Symposium on Habitat Hydraulics, Vol. B, Quebec City, Canada, pp. 215-227.

Allouche, O., Tsoar, A. and Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43, 1223-1232.

Anderson, R.P., Lew, D. and Peterson, A.T. (2003). Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, 162, 211-232.

Araújo, M.B., Pearson, R.G., Thuiller, W. and Erhard, M. (2005). Validation of species-climate impact models under climate change. *Global Change Biology*, 11, 1504-1513.

Araújo, M.B. and Rahbek, C. (2006). How does climate change affect biodiversity? *Science*, 313, 1396-1397.

Armour, C.L. and Taylor, J.G. (1991). Evaluation of the Instream Flow Incremental Methodology by United-States Fish and Wildlife Service Field Users. *Fisheries*, 16, 36-43.

Assilian, S. (1974). Artificial intelligence in the control of real dynamical systems. Ph.D. Thesis, London University, London, Great Britain.

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200, 1-19.

Austin, M.P. (1985). Continuum Concept, Ordination Methods, and Niche Theory. *Annual Review of Ecology and Systematics*, 16, 39-61.

Austin, M.P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157, 101-118.

Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D. and Luoto, M. (2006). Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling*, 199, 197-216.

Austin, M.P., Cunningham, R.B. and Fleming, P.M. (1984). New Approaches to Direct Gradient Analysis Using Environmental Scalars and Statistical Curve-Fitting Procedures. *Vegetatio*, 55, 11-27.

Babuška, R. and Verbruggen, H. (1997). Constructing fuzzy models by product space clustering. In: Hellendoorn, H. and Driankov, D. (Eds.), Fuzzy model identification: selected approaches. Springer, Berlin, Germany, pp. 53-90.

Bailey, S.A., Haines-Young, R.H. and Watkins, C. (2002). Species presence in fragmented landscapes: modelling of species requirements at the national level. *Biological Conservation*, 108, 307-316.

Bain, M. (1995). Habitat at the local scale: Multivariate patterns for stream fishes. *Bulletin Français de la Pêche et de la Pisciculture*, 337/338/339, 165-177.

Baltz, D.M. and Moyle, P.B. (1984). Segregation by Species and Size Classes of Rainbow-Trout, Salmo-Gairdneri, and Sacramento Sucker, Catostomus-Occidentalis, in 3 California Streams. *Environmental Biology of Fishes*, 10, 101-110.

Barry, S. and Elith, J. (2006). Error and uncertainty in habitat models. *Journal of Applied Ecology*, 43, 413-423.

Beebe, J.T. (1996). Fluid speed variability and the importance to managing fish habitat in rivers. *Regulated Rivers-Research & Management*, 12, 63-79.

Belconsulting (2006). Waterbeheerplan voor het deelbekken van de Zwalm. Belconsulting nv, Tielt, Belgium (in Dutch).

Bengtsson, J. (1988). Autecological Studies on *Baetis rhodani* (Pict.) (Ephemeroptera). *Archiv Fur Hydrobiologie*, 111, 615-623.

Berg, A., Gardenfors, U. and von Proschwitz, T. (2004). Logistic regression models for predicting occurrence of terrestrial molluscs in southern Sweden - importance of environmental data quality and model complexity. *Ecography*, 27, 83-93.

Bockelmann, B.N., Fenrich, E.K., Lin, B. and Falconer, R.A. (2004). Development of an ecohydraulics model for stream and river restoration. *Ecological Engineering*, 22, 227-235.

Bodenhofer, U. and Bauer, P. (2005). Interpretability of linguistic variables: A formal account. *Kybernetika*, 41, 227-248.

Booker, D.J. (2003). Hydraulic modelling of fish habitat in urban rivers during high flows. *Hydrological Processes*, 17, 577-599.

Booker, D.J., Dunbar, M.J. and Ibbotson, A. (2004). Predicting juvenile salmonid drift-feeding habitat quality using a three-dimensional hydraulic-bioenergetic model. *Ecological Modelling*, 177, 157-177.

Borsuk, M.E., Stow, C.A. and Reckhow, K.H. (2004). A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*, 173, 219-239.

Bovee, K.D. (1978). Probability of use criteria for the family Salmonidae. Instream Flow Information Paper 4. FWS/OBS-78/07, U.S. Fish and Wildlife Service, Washington, DC.

Bovee, K.D. (1982). A guide to stream habitat analysis using the Instream Flow Incremental Methodology. Instream Flow Information Paper 12, Biological Report 82, US Fish and Wildlife Service, Fort Collins, Colorado, 248 pp.

Bovee, K.D. (1986). Development and evaluation of habitat suitability criteria for use in the instream flow incremental methodology. Instream Flow Information Paper 21, Biological Report 86, US Fish and Wildlife Service, Washington, DC, 235 pp.

Bovee, K.D., Lamb, B.L., Bartholow, J.M., Stalnaker, C.B., Taylor, J. and Henriksen, J. (1998). Stream Habitat Analysis Using the Instream Flow Incremental Methodology. Information and Technology Report USGS/BRD/ITR-1998-0004, Fort Collins, Colorado, 130 pp.

Boyce, M.S., Vernier, P.R., Nielsen, S.E. and Schmiegelow, F.K.A. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157, 281-300.

Bremset, G. and Berg, O.K. (1999). Three-dimensional microhabitat use by young pool-dwelling Atlantic salmon and brown trout. *Animal Behaviour*, 58, 1047-1059.

Brosse, S. and Lek, S. (2000). Modelling roach (Rutilus rutilus) microhabitat using linear and nonlinear techniques. *Freshwater Biology*, 44, 441-452.

Brotons, L., Thuiller, W., Araújo, M.B. and Hirzel, A.H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27, 437-448.

Brugnach, M., Bolte, J. and Bradshaw, G.A. (2003). Determining the significance of threshold values uncertainty in rule-based classification models. *Ecological Modelling*, 160, 63-76.

Brugnach, M., Tagg, A., Keil, F. and de Lange, W.J. (2007). Uncertainty matters: Computer models at the science-policy interface. *Water Resources Management*, 21, 1075-1090.

Buckland, S.T. and Elston, D.A. (1993). Empirical-Models for the Spatial-Distribution of Wildlife. *Journal of Applied Ecology*, 30, 478-495.

Buckley, J.J. (1993). Sugeno Type Controllers Are Universal Controllers. *Fuzzy Sets and Systems*, 53, 299-303.

Bulluck, L., Fleishman, E., Betrus, C. and Blair, R. (2006). Spatial and temporal variations in species occurrence rate affect the accuracy of occurrence models. *Global Ecology and Biogeography*, 15, 27-38.

Burgman, M.A., Breininger, D.R., Duncan, B.W. and Ferson, S. (2001). Setting reliability bounds on habitat suitability indices. *Ecological Applications*, 11, 70-78.

Cadenasso, M.L., Pickett, S.T.A., Weathers, K.C., Bell, S.S., Benning, T.L., Carreiro, M.M. and Dawson, T.E. (2003a). An interdisciplinary and synthetic approach to ecological boundaries. *Bioscience*, 53, 717-722.

Cadenasso, M.L., Pickett, S.T.A., Weathers, K.C. and Jones, C.G. (2003b). A framework for a theory of ecological boundaries. *Bioscience*, 53, 750-758.

Campello, R.J.G.B. and do Amaral, W.C. (2006). Hierarchical fuzzy relational models: Linguistic interpretation and universal approximation. *Ieee Transactions on Fuzzy Systems*, 14, 446-453.

Cantor, S.B., Sun, C.C., Tortolero-Luna, G., Richards-Kortum, R. and Follen, M. (1999). A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. *Journal of Clinical Epidemiology*, 52, 885-892.

Capra, H., Breil, P. and Souchon, Y. (1995). A New Tool to Interpret Magnitude and Duration of Fish Habitat Variations. *Regulated Rivers-Research & Management*, 10, 281-289.

Casillas, J., Cordón, O., Herrera, F. and Magdalena, L. (2003a). Accuracy improvements in linguistic fuzzy modelling. Vol. 129 of Studies in Fuzziness and Soft Computing. Springer Verlag, Heidelberg, 385 pp.

Casillas, J., Cordón, O., Herrera, F. and Magdalena, L. (2003b). Interpretability issues in fuzzy modelling. Vol. 128 of Studies in Fuzziness and Soft Computing, Springer-Verlag, Heidelberg, 643 pp.

Chen, Q.W. and Mynett, A.E. (2003). Integration of data mining techniques and heuristic knowledge in fuzzy logic modelling of eutrophication in Taihu Lake. *Ecological Modelling*, 162, 55-67.

Clarke, R.T., Wright, J.F. and Furse, M.T. (2003). RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling*, 160, 219-233.

Clements, F.E. (1916). Plant succession: an analysis of the development of vegetation. Carnegie Institution of Washington, Publication 242, Washington D.C., 512 pp.

Cochnauer, T. and Elms-Cockrum, T. (1986). Probability-of-use curves for selected Idaho fish species. Project performance report, Idaho Department of Fish and Game, Boise, Idaho, Project F-71-R-10.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Collier, K.J. (1993). Flow preferences of aquatic invertebrates in the Tongariro River. Department of Conservation, Science and Research Series 60, Wellington, New Zealand.

Cramer, J.S. (2003). Logit models: from economics and other fields. Cambridge University Press, 184 pp.

Cummins, K.W. (1973). Trophic Relations of Aquatic Insects. *Annual Review of Entomology*, 18, 183-206.

D'heyere, T., Goethals, P.L.M. and De Pauw, N. (2006). Genetic algorithms for optimisation of predictive ecosystems models based on decision trees and neural networks. *Ecological Modelling*, 195, 20-29.

D'heygere, T., Goethals, P.L.M. and De Pauw, N. (2003). Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic macroinvertebrates. *Ecological Modelling*, 160, 291-300.

Dahl, J. (1998). The impact of vertebrate and invertebrate predators on a stream benthic community. *Oecologia*, 117, 217-226.

De Cooman, W., Peeters, B., Theuns, I., Vos, G., Lammens, S., Debbaudt, W., Timmerman, G., Meers, B., Van Erdeghem, M., Van Wauwe, P., Callebaut, R., Barrez, I., Van den Broeck, S., Emery, J., Van Volsem, S., Bursens, K., Van Hoof, K., D'heygere, T., Soetaert, H., Martens, K., Baten, I., Goris, M., Haustraete, K., Breine, J., Van Thuyne, G., Belpaire, C. and Smis, A. (2007). Milieurapport Vlaanderen, MIRA Achtergrondddocument 2007, Kwaliteit oppervlaktewater. Vlaamse Milieumaatschappij: Aalst, Belgium, 121 pp. (in Dutch).

De Pauw, N. and Vanhooren, G. (1983). Method for biological quality assessment of watercourses in Belgium. *Hydrobiologia*, 100, 153-168.

Dedecker, A.P., Goethals, P.L.M., D'Heygere, T., Gevrey, M., Lek, S. and De Pauw, N. (2005). Application of artificial neural network models to analyse the relationships between Gammarus pulex L. (Crustacea, Amphipoda) and river characteristics. *Environmental Monitoring and Assessment*, 111, 223-241.

Dedecker, A.P., Goethals, P.L.M., Gabriels, W. and De Pauw, N. (2004). Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). *Ecological Modelling*, 174, 161-173.

Dominguez-Dominguez, O., Martinez-Meyer, E., Zambrano, L. and De Leon, G.P.P. (2006). Using ecological-niche Modeling as a conservation tool for freshwater species: Live-bearing fishes in central Mexico. *Conservation Biology*, 20, 1730-1739.

Džeroski, S. (2001). Applications of symbolic machine learning to ecological modelling. *Ecological Modelling*, 146, 263-273.

EAWAG (2002). Fishereiliches Gutachten über die Aarebaggerung in Thun. Forschungsanstalt des ETH-Bereichs, EAWAG (Fish Federal Institute for Environmental Science and Technology), Kastanienbaum, Switzerland, 201 pp. (in German).

Elith, J., Burgman, M.A. and Regan, H.M. (2002). Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling*, 157, 313-329.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. and Zimmermann, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129-151.

Elliott, C.R.N., Dunbar, M.J., Gowing, I. and Acreman, M.C. (1999). A habitat assessment approach to the management of groundwater dominated rivers. *Hydrological Processes*, 13, 459-475.

Elmore, K.L., Weiss, S.J. and Banacos, P.C. (2003). Operational ensemble cloud model forecasts: Some preliminary results. *Weather and Forecasting*, 18, 953-964.

Espinosa, J. and Vandewalle, J. (2000). Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm. *Ieee Transactions on Fuzzy Systems*, 8, 591-600.

EU (2000). Directive of the European Parliament and of the council 2000/60/EC establishing a framework for community action in the field of water policy, European Union. The European Parliament, The Council, PE-CONS 3639/1/00 REV 1 EN.

Fabricus, E. and Gustafson, K.J. (1955). Observation of the spawning behaviour of the grayling, *Thymallus thymallus*. *Report of the Institute of Freshwater Research, Drottningholm*, 36, 75-103.

Fielding, A.H. and Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24, 38-49.

Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J. and Dunn, R.R. (2007). The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography*, 16, 24-33.

Fjellheim, A. (1996). Distribution of benthic invertebrates in relation to stream flow characteristics in a Norwegian river. *Regulated Rivers-Research & Management*, 12, 263-271.

Fogel, D.B. and Ghozeil, A. (1997). A note on representations and variation operators. *Ieee Transactions on Evolutionary Computation*, 1, 159-161.

Forbes, A.D. (1995). Classification-Algorithm Evaluation - 5 Performance-Measures Based on Confusion Matrices. *Journal of Clinical Monitoring*, 11, 189-206.

Gabriels, W., Goethals, P.L.M., Dedecker, A.P., Lek, S. and De Pauw, N. (2007). Analysis of macrobenthic communities in Flanders, Belgium, using a stepwise input variable selection procedure with artificial neural networks. *Aquatic Ecology*, 41, 427-441.

Geman, S., Bienenstock, E. and Doursat, R. (1992). Neural Networks and the Bias Variance Dilemma. *Neural Computation*, 4, 1-58.

Gevrey, M., Dimopoulos, L. and Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160, 249-264.

Ghanem, A., Steffler, P., Hicks, F. and Katopodis, C. (1996). Two-dimensional hydraulic simulation of physical habitat conditions in flowing streams. *Regulated Rivers-Research & Management*, 12, 185-200.

Ginot, V. (1995). EVHA, a Windows software for fish habitat assessment in streams. *Bulletin Français de la Pêche et de la Pisciculture*, 337, 303-308.

Glas, A.S., Lijmer, J.G., Prins, M.H., Bonsel, G.J. and Bossuyt, P.M.M. (2003). The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology*, 56, 1129-1135.

Gleason, H.A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club*, 53, 7-26.

Goethals, P.L.M., Dedecker, A.P., Gabriels, W., Lek, S. and De Pauw, N. (2007). Applications of artificial neural networks predicting macroinvertebrates in freshwaters. *Aquatic Ecology*, 41, 491-508.

Gönczi, A.P. (1989). A study of physical parameters at the spawning sites of the European grayling (*Thymallus thymallus* L.). *Regulated Rivers: Research and Management*, 3, 221-224.

Gore, J.A., Layzer, J.B. and Mead, J. (2001). Macroinvertebrate instream flow studies after 20 years: A role in stream management and restoration. *Regulated Rivers-Research & Management*, 17, 527-542.

Gore, J.A. and Nestler, J.M. (1988). Instream flow studies in perspective. *Regulated Rivers: Research and Management*, 2, 93-101.

Gozlan, R.E., Mastrorillo, S., Copp, G.H. and Lek, S. (1999). Predicting the structure and diversity of young-of-the-year fish assemblages in large rivers. *Freshwater Biology*, 41, 809-820.

Greenberg, L., Svendsen, P. and Harby, A. (1996). Availability of microhabitats and their use by brown trout (Salmo trutta) and grayling (Thymallus thymallus) in the River Vojman, Sweden. *Regulated Rivers-Research & Management*, 12, 287-303.

Guillaume, S. (2001). Designing fuzzy inference systems from data: An interpretability-oriented review. *Ieee Transactions on Fuzzy Systems*, 9, 426-443.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A. and Zimmermann, N.E. (2006a). Using niche-based models to improve the sampling of rare species. *Conservation Biology*, 20, 501-511.

Guisan, A., Graham, C.H., Elith, J. and Huettmann, F. (2007). Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, 13, 332-340.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. and Hastie, T. (2006b). Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, 43, 386-392.

Guisan, A. and Theurillat, J.P. (2000). Equilibrium modeling of alpine plant distribution: how far can we go? *Phytocoenologia*, 30, 353-384.

Guisan, A., Theurillat, J.P. and Kienast, F. (1998). Predicting the potential distribution of plant species in an Alpine environment. *Journal of Vegetation Science*, 9, 65-74.

Guisan, A. and Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993-1009.

Guisan, A., Weiss, S.B. and Weiss, A.D. (1999). GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, 143, 107-122.

Guisan, A. and Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147-186.

Harrell, F.E., Lee, K.L. and Mark, D.B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15, 361-387.

Hastie, T., Tibshirani, R. and Friedman, J.H. (2001). The elements of statistical learning: data mining, inference, and prediction. Springer-Verlag, 536 pp.

Hayes, J.W. and Jowett, I.G. (1994). Microhabitat models of large drift-feeding brown trout in three New Zealand rivers. *North American Journal of Fisheries Management*, 14, 710-725.

Heggenes, J. (1996). Habitat selection by brown trout (Salmo trutta) and young Atlantic salmon (S-salar) in streams: Static and dynamic hydraulic modelling. *Regulated Rivers-Research & Management*, 12, 155-169.

Heikkinen, R.K., Luoto, M., Kuussaari, M. and Toivonen, T. (2007). Modelling the spatial distribution of a threatened butterfly: Impacts of scale and statistical technique. *Landscape and Urban Planning*, 79, 347-357.

Hernandez, P.A., Graham, C.H., Master, L.L. and Albert, D.L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29, 773-785.

Hildrew, A.G. (1992). Food webs and species interactions. In: Calow, P. and Petts, G.E. (Eds.), The rivers handbook: hydrological and ecological principles. Volume 1. Blackwell Scientific Publications, Boston, MA, USA, pp. 309-329.

Hill, D.M. and Hauser, G.E. (1985). The effects of proposed water supply withdrawals on fish habitat in the Piney River. Tennessee Valley Authority, Office of Natural Resources and Economic Development, TVA/ONRED/AWR-86/12.

Hirzel, A. and Guisan, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, 157, 331-341.

Hirzel, A.H., Helfer, V. and Metral, F. (2001). Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, 145, 111-121.

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C. and Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199, 142-152.

Hoehler, F.K. (2000). Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *Journal of Clinical Epidemiology*, 53, 499-503.

Holm, C.F., Armstrong, J.D. and Gilvear, D.J. (2001). Investigating a major assumption of predictive instream habitat models: is water velocity preference of juvenile Atlantic salmon independent of discharge? *Journal of Fish Biology*, 59, 1653-1666.

Holt, R.D. (2003). On the evolutionary ecology of species' ranges. *Evolutionary Ecology Research*, 5, 159-178.

Hudson, H.R., Byrom, A.E. and Chadderton, W.L. (2003). A critique of IFIM - instream habitat simulation in the New Zealand context. Science for Conservation 231, Department of Conservation, Wellington, New Zealand, 69 pp.

Huhta, A., Muotka, T., Juntunen, A. and Yrjonen, M. (1999). Behavioural interactions in stream food webs: the case of drift-feeding fish, predatory invertebrates and grazing mayflies. *Journal of Animal Ecology*, 68, 917-927.

Hüllermeier, E. (2005). Fuzzy methods in machine learning and data mining: Status and prospects. *Fuzzy Sets and Systems*, 156, 387-406.

Hutchinson, G.E. (1957). Population Studies - Animal Ecology and Demography - Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415-427.

Huusko, A. and Yrjana, T. (1997). Effects of instream enhancement structures on brown trout, *Salmo trutta* L., habitat availability in a channelized boreal river: a PHABSIM approach. *Fisheries Management and Ecology*, 4, 453-466.

IBN (1984). Biological water quality: determination of the biotic index based on aquatic macroinvertebrates, NBN T92-402. Institut Belge de Normalisation, Brussel, Belgium.

Jaarsma, N.G., Bergman, M., Schulze, F.H. and de Vaate, A.B. (2007). Macroinvertebrates in a dynamic river environment: analysis of time series from artificial substrates, using a 'white box' neural network modelling method. *Aquatic Ecology*, 41, 413-425.

Jiménez-Valverde, A. and Lobo, J.M. (2006). The ghost of unbalanced species distribution data in geographical model predictions. *Diversity and Distributions*, 12, 521-524.

Jiménez-Valverde, A. and Lobo, J.M. (2007). Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica-International Journal of Ecology*, 31, 361-369.

Jin, Y. (2003). Advanced fuzzy systems design and applications. Vol. 112 of Studies in Fuzziness and Soft Computing. Springer Verlag, Heidelberg, Germany, 269 pp.

Johnson, C.J., Seip, D.R. and Boyce, M.S. (2004). A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. *Journal of Applied Ecology*, 41, 238-251.

Jorde, K., Schneider, M. and Zöllner, F. (2000). Analysis of instream habitat quality - preference functions and fuzzy models. In: Wang, Z.Y. and Hu, S.-X. (Eds.), Stochastic Hydraulics 2000. Balkema, Rotterdam, pp. 671-680.

Jørgensen, S.E. (1999). State-of-the-art of ecological modelling with emphasis on development of structural dynamic models. *Ecological Modelling*, 120, 75-96.

Jørgensen, S.E. (2000). 25 years of ecological modelling by Ecological Modelling. *Ecological Modelling*, 126, 95-99.

Jørgensen, S.E. (2005). Ecological Modelling: editorial overview 2000-2005. *Ecological Modelling*, 188, 137-144.

Jowett, I. (1996). RHYHABSIM, river hydraulics and habitat simulation, computer manual. National Institute of Water and Atmospheric Research (NIWA) report, Hamilton, 50 pp.

Jowett, I.G. (1997). Instream flow methods: A comparison of approaches. *Regulated Rivers-Research & Management*, 13, 115-127.

Jowett, I.G. and Davey, A.J.H. (2007). A comparison of composite habitat suitability indices and generalized additive models of invertebrate abundance and fish presence-habitat availability. *Transactions of the American Fisheries Society*, 136, 428-444.

Jowett, I.G. and Richardson, J. (1990). Microhabitat Preferences of Benthic Invertebrates in a New-Zealand River and the Development of in-Stream Flow-Habitat Models for Deleatidium Spp. *New Zealand Journal of Marine and Freshwater Research*, 24, 19-30.

Jowett, I.G., Richardson, J., Biggs, B.J.F., Hickey, C.W. and Quinn, J.M. (1991). Microhabitat Preferences of Benthic Invertebrates and the Development of Generalized Deleatidium Spp Habitat Suitability Curves, Applied to 4 New-Zealand Rivers. *New Zealand Journal of Marine and Freshwater Research*, 25, 187-199.

Jowett, I.G., Rowe, D. and West, D.W. (1996). Fishery flow requirements of the Tongariro River. Report ELE303, National Institute of Water and Atmospheric Research, Hamilton, New Zealand.

Jungwirth, M., Muhar, S. and Schmutz, S. (2000). Fundamentals of fish ecological integrity and their relation to the extended serial discontinuity concept. *Hydrobiologia*, 422, 85-97.

Kampichler, C., Barthel, J. and Wieland, R. (2000). Species density of foliage-dwelling spiders in field margins: a simple, fuzzy rule-based model. *Ecological Modelling*, 129, 87-99.

Karr, J.R. (1991). Biological Integrity - a Long-Neglected Aspect of Water-Resource Management. *Ecological Applications*, 1, 66-84.

Katopodis, C. (2003). Case studies of instream flow modelling for fish habitat in Canadian prairie rivers. *Canadian Water Resources Journal*, 28, 199-216.

Kearns, B.L., Karr, J.R. and Ahlstedt, S.A. (1992). Aquatic invertebrate assemblages: spatial and temporal differences among sampling protocols. *Journal of the North American Benthological Society*, 11, 377-390.

Kelly, D.W., Dick, J.T.A. and Montgomery, I. (2002). Predation on mayfly nymph, Baetis rhodani, by native and introduced Gammarus: direct effects and the facilitation of predation by salmonids. *Freshwater Biology*, 47, 1257-1268.

Kemp, P.S., Gilvear, D.J. and Armstrong, J.D. (2003). Do juvenile Atlantic salmon parr track local changes in water velocity? *River Research and Applications*, 19, 569-575.

King, J., Brown, C. and Sabet, H. (2003). A scenario-based holistic approach to environmental flow assessments for rivers. *River Research and Applications*, 19, 619-639.

Korzukhin, M.D., TerMikaelian, M.T. and Wagner, R.G. (1996). Process versus empirical models: Which approach for forest ecosystem management? *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 26, 879-887.

Kruse, R., Gebhardt, J. and Klawonn, F. (1994). Foundations of fuzzy systems. John Wiley, Chichester, UK, 278 pp.

Lambert, T.R. and Hanson, D.F. (1989). Development of habitat suitability criteria for trout in small streams. *Regulated Rivers: Research and Management*, 3, 291-303.

Lamouroux, N., Capra, H. and Pouilly, M. (1998). Predicting habitat suitability for lotic fish: Linking statistical hydraulic models with multivariate habitat use models. *Regulated Rivers-Research & Management*, 14, 1-11.

Lamouroux, N., Souchon, Y. and Herouin, E. (1995). Predicting Velocity Frequency-Distributions in Stream Reaches. *Water Resources Research*, 31, 2367-2375.

Lantz, C.A. and Nebenzahl, E. (1996). Behavior and interpretation of the kappa statistic: Resolution of the two paradoxes. *Journal of Clinical Epidemiology*, 49, 431-434.

Leathwick, J.R. and Whitehead, D. (2001). Soil and atmospheric water deficits and the distribution of New Zealand's indigenous tree species. *Functional Ecology*, 15, 233-242.

Leclerc, M. (1983). Marquage et repérage radiotélémétriques de 71 especes de poisson du territoire central du projet archipel entre octobre 1982 et août 1983, synthèse des résultats. In: Archipel de Montréal, Rapport Technique, Poisson 5, Québec, Canada, pp. 152-153.

Legendre, P. (1993). Spatial Autocorrelation - Trouble or New Paradigm. *Ecology*, 74, 1659-1673.

Lek, S. and Guégan, J.F. (1999). Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, 120, 65-73.

Levins, R. (1966). Strategy of Model Building in Population Biology. *American Scientist*, 54, 421-431.

Liu, C.R., Berry, P.M., Dawson, T.P. and Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28, 385-393.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G. and Williams, P.H. (2003). Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology*, 17, 1591-1600.

Luck, G.W. (2002). The habitat requirements of the rufous treecreeper (Climacteris rufa). 2. Validating predictive habitat models. *Biological Conservation*, 105, 395-403.

Ludwig, D. (2001). The era of management is over. *Ecosystems*, 4, 758-764.

Luoto, M., Poyry, J., Heikkinen, R.K. and Saarinen, K. (2005). Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and Biogeography*, 14, 575-584.

Lütolf, M., Kienast, F. and Guisan, A. (2006). The ghost of past species occurrence: improving species distribution models for presence-only data. *Journal of Applied Ecology*, 43, 802-815.

Mac Nally, R., Fleishman, E., Fay, J.P. and Murphy, D.D. (2003). Modelling butterfly species richness using mesoscale environmental variables: model construction and validation for mountain ranges in the Great Basin of western North America. *Biological Conservation*, 110, 21-31.

MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. and Franklin, A.B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84, 2200-2207.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. and Langtimm, C.A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248-2255.

Mackenzie, D.I. and Royle, J.A. (2005). Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, 42, 1105-1114.

Maclure, M. and Willett, W.C. (1987). Misinterpretation and Misuse of the Kappa-Statistic. *American Journal of Epidemiology*, 126, 161-169.

Maddock, I. (1999). The importance of physical habitat assessment for evaluating river health. *Freshwater Biology*, 41, 373-391.

Maggini, R., Lehmann, A., Zimmermann, N.E. and Guisan, A. (2006). Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography*, 33, 1729-1749.

MakiPetays, A., Muotka, T., Huusko, A., Tikkanen, P. and Kreivi, P. (1997). Seasonal changes in habitat use and preference by juvenile brown trout, Salmo trutta, in a northern boreal river. *Canadian Journal of Fisheries and Aquatic Sciences*, 54, 520-530.

Malanson, G.P. (1997). Simulated responses to hypothetical fundamental niches. *Journal of Vegetation Science*, 8, 307-316.

Mamdani, E. (1974). Application of fuzzy algorithms for control of a simple dynamic plant. *Proceedings IEE*, 121, 1585-1588.

Manel, S., Dias, J.M., Buckton, S.T. and Ormerod, S.J. (1999a). Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, 36, 734-747.

Manel, S., Dias, J.M. and Ormerod, S.J. (1999b). Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, 120, 337-347.

Manel, S., Williams, H.C. and Ormerod, S.J. (2001). Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, 38, 921-931.

McBride, J.L. and Ebert, E.E. (2000). Verification of quantitative precipitation forecasts from operational numerical weather prediction models over Australia. *Weather and Forecasting*, 15, 103-121.

McNyset, K.M. and Blackburn, J.K. (2006). Does GARP really fail miserably? A response to Stockman et al. (2006). *Diversity and Distributions*, 12, 782-786.

McPherson, J.M. and Jetz, W. (2007). Effects of species' ecology on the accuracy of distribution models. *Ecography*, 30, 135-151.

McPherson, J.M., Jetz, W. and Rogers, D.J. (2004). The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, 41, 811-823.

Meentemeyer, R.K., Moody, A. and Franklin, J. (2001). Landscape-scale patterns of shrub-species abundance in California chaparral - The role of topographically mediated resource gradients. *Plant Ecology*, 156, 19-41.

Metternicht, G. (2001). Assessing temporal and spatial changes of salinity using fuzzy logic, remote sensing and GIS. Foundations of an expert system. *Ecological Modelling*, 144, 163-179.

Meynard, C.N. and Quinn, J.F. (2007). Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, 34, 1455-1469.

Michalewicz, Z. and Fogel, D.B. (2000). How to solve it: modern heuristics. Springer-Verlag, Berlin, Heidelberg, 467 pp.

Milhous, R.T., Updike, M.A. and Schneider, D.M. (1989). Physical habitat simulation system reference manual - version II. . Instream Flow Information Paper 26, United States Fish and Wildlife Service, Fort Collins, Colorado, 403 pp.

Miller, M.E., Hui, S.L. and Tierney, W.M. (1991). Validation Techniques for Logistic-Regression Models. *Statistics in Medicine*, 10, 1213-1226.

Minshall, G.W. (1984). Aquatic insect-substratum relationships. In: Resh, V.H. and Rosenberg, D.M. (Eds.), The ecology of aquatic insects. Praeger, New York, pp. 358-400.

Mobes-Hansen, B. and Waringer, J.A. (1998). The influence of hydraulic stress on microdistribution patterns of zoobenthos in a sandstone brook (Weidlingbach, Lower Austria). *International Review of Hydrobiology*, 83, 381-396.

Moir, H.J., Gibbins, C.N., Soulsby, C. and Youngson, A.F. (2005). PHABSIM modelling of Atlantic salmon spawning habitat in an upland stream: Testing the influence of habitat suitability indices on model output. *River Research and Applications*, 21, 1021-1034.

Mouton, A., Meixner, H., Goethals, P.L.M., De Pauw, N. and Mader, H. (2007a). Concept and application of the usable volume for modelling the physical habitat of riverine organisms. *River Research and Applications*, 23, 545-558.

Mouton, A.M., Schneider, M., Depestele, J., Goethals, P.L.M. and De Pauw, N. (2007b). Fish habitat modelling as a tool for river management. *Ecological Engineering*, 29, 305-315.

Mouton, A.M., Schneider, M., Kopecki, I., Goethals, P.L.M. and De Pauw, N. (2006). Application of MesoCASiMiR: Assessment of *Baetis rhodani* habitat suitability. In: Voinov, A., Jakeman, A.J. and Rizzoli, A.E. (Eds.), Proceedings of the iEMSs Third Biennial Meeting: "Summit on Environmental Modelling and Software", International Modelling and Software Society, Burlington, USA, CD-ROM publication.

Mouton, A.M., Schneider, M., Kopecki, I., Goethals, P.L.M. and De Pauw, N. (submitted). Optimisation of physical habitat assessment for benthic fish and macroinvertebrates in lowland rivers. *Journal of Environmental Monitoring*, submitted.

Mouton, A.M., Van der Most, H., Jeuken, A., Goethals, P.L.M. and De Pauw, N. (2008). Evaluation of river basin restoration options by the application of the Water Framework Directive Explorer in the Zwalm River basin (Flanders, Belgium). *River Research and Applications*, 23, 1-16.

Moyle, P.B. and Baltz, D.M. (1985). Microhabitat Use by an Assemblage of California Stream Fishes - Developing Criteria for Instream Flow Determinations. *Transactions of the American Fisheries Society*, 114, 695-704.

Murphy, A.H. and Winkler, R.L. (1987). A General Framework for Forecast Verification. *Monthly Weather Review*, 115, 1330-1338.

Murphy, A.H. and Winkler, R.L. (1992). Diagnostic Verification of Probability Forecasts. *International Journal of Forecasting*, 7, 435-455.

Nelson, L.D. and Cicchetti, D.V. (1995). Assessment of emotional functioning in brain-impaired individuals. *Psychological Assessment*, 7, 404-413.

Nestler, J. and Sutton, V.K. (2000). Describing scales of features in river channels using fractal geometry concepts. *Regulated Rivers-Research & Management*, 16, 1-22.

Newson, M.D. and Newson, C.L. (2000). Geomorphology, ecology and river channel habitat: mesoscale approaches to basin-scale challenges. *Progress in Physical Geography*, 24, 195-217.

Nielsen, S.E., Johnson, C.J., Heard, D.C. and Boyce, M.S. (2005). Can models of presence-absence be used to scale abundance? - Two case studies considering extremes in life history. *Ecography*, 28, 197-208.

Norris, R.H. and Norris, K.R. (1995). The need for biological assessment of water quality: Australian perspective. *Australian Journal of Ecology*, 20, 1-6.

Nykänen, M. and Huusko, A. (2002). Suitability criteria for spawning habitat of riverine European grayling. *Journal of Fish Biology*, 60, 1351-1354.

Nykänen, M., Huusko, A. and Lahti, M. (2004). Changes in movement, range and habitat preferences of adult grayling from late summer to early winter. *Journal of Fish Biology*, 64, 1386-1398.

Nykänen, M., Huusko, A. and Maki-Petays, A. (2001). Seasonal changes in the habitat use and movements of adult European grayling in a large subarctic river. *Journal of Fish Biology*, 58, 506-519.

Olden, J.D. and Jackson, D.A. (2002). A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology*, 47, 1976-1995.

Omlin, M. and Reichert, P. (1999). A comparison of techniques for the estimation of model prediction uncertainty. *Ecological Modelling*, 115, 45-59.

Osborne, P.E. and Suarez-Seoane, S. (2002). Should data be partitioned spatially before building large-scale distribution models? *Ecological Modelling*, 157, 249-259.

Pahl-Wostl, C., Lindenschmidt, K.E., Janssen, J.A.E.B., Filatova, T., Mouton, A., Holtz, G., Van der Keur, P. and Gaber, N. (in press). Complexity and uncertainty: rethinking the modelling activity. In: Jakeman, T., Voinov, A., Rizzoli, A. and Chen, S. (Eds.), Integrated and participatory water resources management.

Palmer, M.A., Bernhardt, E.S., Allan, J.D., Lake, P.S., Alexander, G., Brooks, S., Carr, J., Clayton, S., Dahm, C.N., Shah, J.F., Galat, D.L., Loss, S.G., Goodwin, P., Hart, D.D., Hassett, B., Jenkinson, R., Kondolf, G.M., Lave, R., Meyer, J.L., O'Donnell, T.K., Pagano, L. and Sudduth, E. (2005). Standards for ecologically successful river restoration. *Journal of Applied Ecology*, 42, 208-217.

Parasiewicz, P. (2001). MesoHABSIM: A concept for application of instream flow models in river restoration planning. *Fisheries*, 26, 6-13.

Parasiewicz, P. (2007a). The MesoHABSIM model revisited. *River Research and Applications*, 23, 893-903.

Parasiewicz, P. (2007b). Overcoming the limits of scales. *River Research and Applications*, 23, 891-892.

Parasiewicz, P. and Dunbar, M.J. (2001). Physical habitat modelling for fish - a developing approach. *Archiv für Hydrobiologie Supplement*, 135, 1-30.

Parasiewicz, P. and Walker, J.D. (2007). Comparison of MesoHABSIM with two microhabitat models (PHABSIM and HARPHA). *River Research and Applications*, 23, 904-923.

Parra, J.L., Graham, C.C. and Freile, J.F. (2004). Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography*, 27, 350-360.

Pasternack, G.B., Wang, C.L. and Merz, J.E. (2004). Application of a 2D hydrodynamic model to design of reach-scale spawning gravel replenishment on the Mokelumne River, California. *River Research and Applications*, 20, 205-225.

Pearce, J. and Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133, 225-245.

Pearson, R.G., Dawson, T.P. and Liu, C. (2004). Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, 27, 285-298.

Perfilieva, I. and Kreinovich, V. (2002). A new universal approximation result for fuzzy systems, which reflects CNF-DNF duality. *International Journal of Intelligent Systems*, 17, 1121-1130.

Pert, E.J. and Erman, D.C. (1994). Habitat Use by Adult Rainbow-Trout under Moderate Artificial Fluctuations in Flow. *Transactions of the American Fisheries Society*, 123, 913-923.

Peterson, A.T. and Holt, R.D. (2003). Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. *Ecology Letters*, 6, 774-782.

Peterson, A.T., Papes, M. and Eaton, M. (2007). Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, 30, 550-560.

Petit, S., Chamberlain, D., Haysom, K., Pywell, R., Vickery, J., Warman, L., Allen, D. and Firbank, L. (2003). Knowledge-based models for predicting species occurrence in arable conditions. *Ecography*, 26, 626-640.

Petts, G.E., Nestler, J. and Kennedy, R. (2006). Advancing science for water resources management. *Hydrobiologia*, 565, 277-288.

Pickett, S.T.A., Kolasa, G. and Jones, C.G. (1994). Ecological understanding: The nature of theory and the theory of nature. Academic Press, New York, 206 pp.

Poff, N.L., Allan, J.D., Palmer, M.A., Hart, D.D., Richter, B.D., Arthington, A.H., Rogers, K.H., Meyers, J.L. and Stanford, J.A. (2003). River flows and water wars: emerging science for environmental decision making. *Frontiers in Ecology and the Environment*, 1, 298-306.

Pulliam, H.R. (2000). On the relationship between niche and distribution. *Ecology Letters*, 3, 349-361.

Raleigh, R.F., Hickman, T., Solomon, R.C. and Nelson, P.C. (1984). Habitat suitability information: rainbow trout. U.S. Fish and Wildlife Service Biological Services Program, FWS/OBS/-82/10.60.

Randin, C.F., Dirnbock, T., Dullinger, S., Zimmermann, N.E., Zappa, M. and Guisan, A. (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography*, 33, 1689-1703.

Regan, H.M., Colyvan, M. and Burgman, M.A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications*, 12, 618-628.

Resh, V.H., Hildrew, A.G., Statzner, B. and Townsend, C.R. (1994). Theoretical Habitat Templets, Species Traits, and Species Richness - a Synthesis of Long-Term Ecological Research on the Upper Rhone River in the Context of Concurrently Developed Ecological Theory. *Freshwater Biology*, 31, 539-554.

Reynoldson, T.B., Bailey, R.C., Day, K.E. and Norris, R.H. (1995). Biological Guidelines for Fresh-Water Sediment Based on Benthic Assessment of Sediment (the Beast) Using a Multivariate Approach for Predicting Biological State. *Australian Journal of Ecology*, 20, 198-219.

Rhoads, B.L., Schwartz, J.S. and Porter, S. (2003). Stream geomorphology, bank vegetation, and three-dimensional habitat hydraulics for fish in midwestern agricultural streams. *Water Resources Research*, 39, 2-13.

Riley, W.D., Ives, M.J., Pawson, M.G. and Maxwell, D.L. (2006). Seasonal variation in habitat use by salmon, Salmo salar, trout, Salmo trutta and grayling, Thymallus thymallus, in a chalk stream. *Fisheries Management and Ecology*, 13, 221-236.

Robertson, M.P., Peter, C.I., Villet, M.H. and Ripley, B.S. (2003). Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling*, 164, 153-167.

Rogers, K.H. (2006). The real river management challenge: Integrating scientists, stakeholders and service agencies. *River Research and Applications*, 22, 269-280.

Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H. and Possingham, H.P. (2006). Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters*, 9, 1136-1145.

Roni, P., Beechie, T.J., Bilby, R.E., Leonetti, F.E., Pollock, M.M. and Pess, G.R. (2002). A review of stream restoration techniques and a hierarchical strategy for prioritizing restoration in Pacific northwest watersheds. *North American Journal of Fisheries Management*, 22, 1-20.

Rosenberg, D.M. and Resh, V.H. (1993). Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York, 488 pp.

Rosenfeld, J. (2003). Assessing the habitat requirements of stream fishes: An overview and evaluation of different approaches. *Transactions of the American Fisheries Society*, 132, 953-968.

Rouget, M., Richardson, D.M., Nel, J.L., Le Maitre, D.C., Egoh, B. and Mgidi, T. (2004). Mapping the potential ranges of major plant invaders in South Africa, Lesotho and Swaziland using climatic suitability. *Diversity and Distributions*, 10, 475-484.

Roux, D.J., Rogers, K.H., Biggs, H.C., Ashton, P.J. and Sergeant, A. (2006). Bridging the science-management divide: Moving from unidirectional knowledge transfer to knowledge interfacing and sharing. *Ecology and Society*, 11, 4.

Rushton, S.P., Ormerod, S.J. and Kerby, G. (2004). New paradigms for modelling species distributions? *Journal of Applied Ecology*, 41, 193-200.

Ruspini, E.H. (1969). A New Approach to Clustering. *Information and Control*, 15, 22-32.

Salski, A. (1992). Fuzzy knowledge-based models in ecological research. *Ecological Modelling*, 63, 103-112.

Saseendran, S.A., Singh, S.V., Rathore, L.S. and Das, S. (2002). Characterization of weekly cumulative rainfall forecasts over meteorological subdivisions of India using a GCM. *Weather and Forecasting*, 17, 832-844.

Schmidt-Kloiber, A. and Nijboer, R.C. (2004). The effect of taxonomic resolution on the assessment of ecological water quality classes. *Hydrobiologia*, 516, 269-283.

Schneider, M. (2001). Habitat- und Abflussmodellierung für Fliessgewässer mit unscharfen Berechnungsansätzen. PhD. Thesis, Institut für Wasserbau, Universität Stuttgart, 158 pp.

Scott, J.M., Heglund, P.J., Samson, F., Haufler, J., Morrison, M., Raphael, M. and Wall, B. (2002). Predicting species occurrences: Issues of accuracy and scale. Island Press, Covelo, California, 868 pp.

Segurado, P. and Araújo, M.B. (2004). An evaluation of methods for modelling species distributions. *Journal of Biogeography*, 31, 1555-1568.

Sekine, M., Imai, T. and Ukita, M. (1997). A model of fish distribution in rivers according to their preference for environmental factors. *Ecological Modelling*, 104, 215-230.

Sempeski, P. and Gaudin, P. (1995). Habitat Selection by Grayling .1. Spawning Habitats. *Journal of Fish Biology*, 47, 256-265.

Shannon, C.E. and Weaver, W. (1963). Mathematical theory of communication. University of Illinois Press, Urbana, Illinois, 144 pp.

Shields, F.D., Knight, S.S. and Cooper, C.M. (1997). Rehabilitation of warmwater stream ecosystems following channel incision. *Ecological Engineering*, 8, 93-116.

Shuler, S.W. and Nehring, R.B. (1994). Using the Physical Habitat Simulation Model to evaluate a stream habitat enhancement project. *Rivers* 4, 175-193.

Silvert, W. (2000). Fuzzy indices of environmental conditions. *Ecological Modelling*, 130, 111-119.

Stalnaker, C. (1995). The instream flow incremental methodology: a primer for IFIM. National Ecology Research Centre, Internal Publication. U.S. Dept. of the interior, Nat. Biol. Service, Fort Collins, Colorado, 45 pp.

Stalnaker, C.B., Bovee, K.D. and Waddle, T.J. (1996). Importance of the temporal aspects of habitat hydraulics to fish population studies. *Regulated Rivers-Research & Management*, 12, 145-153.

Stark, J.D. and Suren, A.M. (2003). Project aqua: environmental study - aquatic ecosystems: macroinvertebrates. National Institute of Water and Atmospheric Research, Report CHC02/59, Christchurch, New Zealand.

Statzner, B., Gore, J.A. and Resh, V.H. (1988). Hydraulic Stream Ecology - Observed Patterns and Potential Applications. *Journal of the North American Benthological Society*, 7, 307-360.

Stewart, G., Anderson, R. and Wohl, E. (2005). Two-dimensional modelling of habitat suitability as a function of discharge on two Colorado rivers. *River Research and Applications*, 21, 1061-1074.

Stockman, A.K., Beamer, D.A. and Bond, J.E. (2006). An evaluation of a GARP model as an approach to predicting the spatial distribution of non-vagile invertebrate species. *Diversity and Distributions*, 12, 81-89.

Stockwell, D.R.B. and Peterson, A.T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148, 1-13.

Strahler, A.N. (1957). Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union*, 38, 913-920.

Strauss, B. and Biedermann, R. (2007). Evaluating temporal and spatial generality: How valid are species-habitat relationship models? *Ecological Modelling*, 204, 104-114.

Strayer, D.L., Power, M.E., Fagan, W.F., Pickett, S.T.A. and Belnap, J. (2003). A classification of ecological boundaries. *Bioscience*, 53, 723-729.

Suchanek, P.M., Kuntz, K.J. and McDonnell, J.P. (1984). The relative abundance, distribution and instream flow relationships of juvenile salmon in the lower Sustina River. Sustina River Aquatic Studies Program Report 7, part 2, Alaska Department of Fish and Game, Anchorage, Alaska.

Teirney, L.D. and Jowett, I.G. (1990). Trout abundance in New Zealand rivers: an assessment by drift diving. Ministry of Agriculture and Fisheries, Freshwater Fisheries Centre, Freshwater Fisheries Report 118, Christchurch, New Zealand.

Termansen, M., McClean, C.J. and Preston, C.D. (2006). The use of genetic algorithms and Bayesian classification to model species distributions. *Ecological Modelling*, 192, 410-424.

Thomas, J.A. and Bovee, K.D. (1993). Application and Testing of a Procedure to Evaluate Transferability of Habitat Suitability Criteria. *Regulated Rivers-Research & Management*, 8, 285-294.

Thuiller, W., Araújo, M.B. and Lavorel, S. (2003). Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, 14, 669-680.

Thuiller, W., Lavorel, S. and Araújo, M.B. (2005). Niche properties and geographical extent as predictors of species sensitivity to climate change. *Global Ecology and Biogeography*, 14, 347-357.

Tikkanen, P., Huhta, A. and Muotka, T. (2000). Determinants of substrate selection in lotic mayfly larvae: is cryptic coloration important? *Archiv Fur Hydrobiologie*, 148, 45-57.

Tikkanen, P., Muotka, T. and Huhta, A. (1996). Fishless-stream mayflies express behavioural flexibility in response to predatory fish. *Animal Behaviour*, 51, 1391-1399.

Townsend, C.R. (1989). The Patch Dynamics Concept of Stream Community Ecology. *Journal of the North American Benthological Society*, 8, 36-50.

Townsend, C.R., Doledec, S. and Scarsbrook, M.R. (1997). Species traits in relation to temporal and spatial heterogeneity in streams: A test of habitat templet theory. *Freshwater Biology*, 37, 367-387.

Tyre, A.J., Possingham, H.P. and Lindenmayer, D.B. (2001). Inferring process from pattern: Can territory occupancy provide information about life history parameters? *Ecological Applications*, 11, 1722-1737.

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. and Possingham, H.P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications*, 13, 1790-1801.

Ussegliopolatera, P. (1994). Theoretical Habitat Templets, Species Traits, and Species Richness - Aquatic Insects in the Upper Rhone River and Its Floodplain. *Freshwater Biology*, 31, 417-437.

Van Broekhoven, E., Adriaenssens, V. and De Baets, B. (2007). Interpretability-preserving genetic optimization of linguistic terms in fuzzy models for fuzzy ordered classification: An ecological case study. *International Journal of Approximate Reasoning*, 44, 65-90.

Van Broekhoven, E., Adriaenssens, V., De Baets, B. and Verdonschot, P.F.M. (2006). Fuzzy rule-based macroinvertebrate habitat suitability models for running waters. *Ecological Modelling*, 198, 71-84.

Van der Most, H., Jeuken, A. and Van Schijndel, S. (2006). WFD-Explorer: A planning kit to support the implementation of the Water Framework Directive. In: Gourbesville, P., Cunge, J., Guinot, V. and Liong, S.-Y. (Eds.), Proceedings of the 7th International Conference in Hydroinformatics, Volume I, Nice, Research Publishing Services, Chennai, India, pp. 8-15.

Vanhouwelingen, J.C. and Lecessie, S. (1990). Predictive Value of Statistical-Models. *Statistics in Medicine*, 9, 1303-1325.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. and Cushing, C.E. (1980). River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences*, 37, 130-137.

Vanreusel, W., Maes, D. and Van Dyck, H. (2007). Transferability of species distribution models: a functional habitat approach for two regionally threatened butterflies. *Conservation Biology*, 21, 201-212.

Vaughan, I.P. and Ormerod, S.J. (2005). The continuing challenges of testing species distribution models. *Journal of Applied Ecology*, 42, 720-730.

Vilizzi, L. and Copp, G.H. (2005). An analysis of 0+barbel (Barbus barbus) response to discharge fluctuations in a flume. *River Research and Applications*, 21, 421-438.

Vilizzi, L., Copp, G.H. and Roussel, J.M. (2004). Assessing variation in suitability curves and electivity profiles in temporal studies of fish habitat use. *River Research and Applications*, 20, 605-618.

Vismara, R., Azzellino, A., Bosi, R., Crosa, G. and Gentili, G. (2001). Habitat suitability curves for brown trout (Salmo trutta fario L.) in the River Adda, Northern Italy: Comparing univariate and multivariate approaches. *Regulated Rivers-Research & Management*, 17, 37-50.

VMM (2007). Water- & waterbodemkwaliteit - Lozingen in het water - Evaluatie saneringsinfrastructuur 2005. Vlaamse Milieumaatschappij, Aalst, Belgium, 128 pp. (in Dutch).

Wagner, R., Dapper, T. and Schmidt, H.H. (2000). The influence of environmental variables on the abundance of aquatic insects: a comparison of ordination and artificial neural networks. *Hydrobiologia*, 422, 143-152.

Waters, B.F. (1976). A methodology for evaluating the effects of different stream flows on salmonid habitat. In: Osborn, J.F. and Allman, C.H. (Eds.), Instream flow needs. American Fisheries Society, Bethesda, Maryland, USA, pp. 254-266.

Welsh, A.H. (1996). Aspects of Statistical Inference. John Wiley & Sons, New York, USA, 452 pp.

Wiens, J.J. and Graham, C.H. (2005). Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology Evolution and Systematics*, 36, 519-539.

Wilson, K.A., Westphal, M.I., Possingham, H.P. and Elith, J. (2005). Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation*, 122, 99-112.

Wolpert, D.H. and Macready, W.G. (1997). No free lunch theorems for optimization. *Ieee Transactions on Evolutionary Computation*, 1, 67-82.

Wood, P.J., Toone, J., Greenwood, M.T. and Armitage, P.D. (2005). The response of four lotic macroinvertebrate taxa to burial by sediments. *Archiv Fur Hydrobiologie*, 163, 145-162.

Woolf, A., Nielsen, C.K., Weber, T. and Gibbs-Kieninger, T.J. (2002). Statewide modeling of bobcat, Lynx rufus, habitat in Illinois, USA. *Biological Conservation*, 104, 191-198.

Wright, J.F., Furse, M.T., Armitage, P.D. and Moss, D. (1993). New Procedures for Identifying Running-Water Sites Subject to Environmental-Stress and for Evaluating Sites for Conservation, Based on the Macroinvertebrate Fauna. *Archiv Fur Hydrobiologie*, 127, 319-326.

Wright, J.F., Moss, D., Armitage, P.D. and Furse, M.T. (1984). A Preliminary Classification of Running-Water Sites in Great-Britain Based on Macroinvertebrate Species and the Prediction of Community Type Using Environmental Data. *Freshwater Biology*, 14, 221-256.

Ying, H., Ding, Y.S., Li, S.K. and Shao, S.H. (1999). Comparison of necessary conditions for typical Takagi-Sugeno and Mamdani fuzzy systems as universal approximators. *IEEE Transactions on Systems Man and Cybernetics Part a-Systems and Humans*, 29, 508-514.

Zadeh, L.A. (1965). Fuzzy Sets. *Information and Control*, 8, 338-353.

Žnidaršic, M., Jakulin, A., Džeroski, S. and Kampichler, C. (2006). Automatic construction of concept hierarchies: The case of foliage-dwelling spiders. *Ecological Modelling*, 191, 144-158.

Summary

For many years, river management across the world has focused on the improvement of the chemical water quality, using instruments such as wastewater treatment plants and enforced effluent standards. Although these measures resulted in a significant improvement of the ecological river quality, many streams still show a poor physical habitat quality and the ecological quality is often stagnating. To meet the aims set by the European Water Framework Directive, small-scale measures such as remeandering, flood plane restoration and fish passages are needed and are now being planned or implemented across Europe. However, efficient allocation of these small scale efforts requires thorough analysis of their possible impact on river ecology. Since habitat suitability models enable this analysis and in particular the identification of the bottlenecks in a river basin, these models are valuable tools for integrated river management.

To allow decision makers to assess the uncertainty associated with the model outputs, the model structure should be straightforwardly interpretable. Fuzzy logic has become an interesting technique to address this issue. It takes into account the inherent uncertainty of ecological variables during inference processing and it enables expressing non-linear relations between ecological variables in a transparent way. Fuzzy systems use linguistic descriptions such as ‘low’, ‘high’ or ‘moderate’ for quantification of variables and use ecological expert knowledge to transform these descriptions into a mathematical framework in which data processing can be performed. Specifically, the expert knowledge is embedded in fuzzy if-then rules such as 'IF the depth is low AND the flow velocity is high THEN the habitat suitability is high for a rheophilic fish species'. Fuzzy models use fuzzy sets to implement the uncertainty in the definition of the linguistic values 'low', 'high' and 'high' of the linguistic variables 'depth', 'flow velocity' and 'habitat suitability'. Since the boundaries of two fuzzy sets are overlapping, an object (e.g. a real value) may partially belong to two consecutive sets and thus a vague linguistic description such as 'the depth is quite low but tending to be moderate' can be incorporated into the model. This approach contrasts markedly with classical set theory, where an object either fully belongs to a set or not.

Fuzzy species distribution models allow incorporation of ecological aspects, such as the ecological gradient theory, into the modelling process. This turns fuzzy modelling into an appropriate technique for species distribution modelling because this approach may reflect the specific characteristics of the ecological problem. However, the main bottleneck in the application of fuzzy logic is the need for ecological expert knowledge. Not only is the formalisation of problem-relevant human expert knowledge often difficult and tedious, but several authors also contest its consistency. Recent research has shown that complementing fuzzy systems by data-driven techniques can solve this knowledge acquisition bottleneck. For example, the induction of fuzzy rule-based models by heuristic search algorithms is often used in the field of fuzzy rule learning. This dissertation aimed to develop and test a data-driven optimisation method for fuzzy habitat suitability models. Specifically, a hill-climbing algorithm was applied to optimise the fuzzy expert rules of the model and the impact of different training criteria on the optimisation results was analysed. The fuzzy sets were optimised based on their entropy, which quantifies the uniformity of the data distribution over the input space. The value of the presented approach for integrated river management was tested on 5 datasets which described ecological case studies in Belgium, Switzerland and New-Zealand.

First, the strengths and weaknesses of an expert knowledge-based approach was analysed by comparing it to a data-driven method in a case study in the Zwalm river (Flanders, Belgium). Based on the available ecological expert knowledge and on biological samples from the Zwalm river basin, habitat models were generated for the mayfly *Baetis rhodani*, which is an indicator species for a good ecological water quality. Both fuzzy expert knowledge-based and data-driven habitat suitability models were developed and the performance of these models was compared. For the data-driven fuzzy models, a hill-climbing optimisation algorithm was applied to derive ecological knowledge from the available data. Data-driven models appeared to outperform expert knowledge-based models significantly, while a step forward model selection procedure indicated that physical habitat variables such as flow velocity, river width and the dominant river bed substrate adequately described the mayfly habitat suitability in the studied area. This study has important implications on the application of expert knowledge in ecological studies, especially if this knowledge is extrapolated to other areas. The results suggest that data-driven models can complement expert-knowledge based approaches and hence improve model reliability.

To develop reliable data-driven models, however, a sound model training and evaluation procedure is needed. A crucial step in these procedures is the assessment of the model performance. The key component of model performance assessment is the applied performance criterion. Therefore, this dissertation reviewed the performance criteria most commonly applied in presence-absence modelling and analysed their function in both the model training and evaluation process. Although estimates of predictive performance have been used widely to assess final model quality, an extended literature review in this dissertation showed that the model training procedure has been mostly based on optimisation of the predictive accuracy of the model. Moreover, a systematic overview was lacking because most analyses of performance criteria have been empirical and only focused on specific aspects of the performance criteria. This dissertation provides such an overview which showed that different performance criteria evaluate a model (or its resulting confusion matrix) differently and that this difference may be explained by the dependency of these criteria on the prevalence of the validation set. Specifically, if models derived from different datasets are being compared, the prevalence of these datasets may affect the value of the performance measures and consequently manipulate the results of the comparison. Similar problems may arise when the performances of a model on a training set and a validation set with different prevalences are compared. This dissertation showed theoretically that these effects only occur if the data are inseparable by an n-dimensional hyperplane, with n the number of input variables. Given this inseparability, different performance criteria focus on different aspects of model performance during model training, such as sensitivity, specificity or predictive accuracy. These findings have important consequences for ecological modelling because ecological data are mostly inseparable due to data noise and the complexity of the studied system. Consequently, it should be very clear which aspect of the model performance is evaluated, and models should be evaluated consistently, that is, independent of, or taking into account, species prevalence.

This dissertation also highlighted the relative importance which the performance criteria give to omission and commission errors as a possible explanation for the differing evaluation scores among performance criteria for the same model. Theoretical analysis revealed that performance criteria may value a perfect model equally, but yet focus on very different aspects of model performance. To guarantee reliable quantification of the model performance, it is suggested that prevalence-independent measures should be preferred, and that at least some of these measures should allow modellers to distinguish between omission and commission errors for both model training and evaluation. The appropriate performance criteria for evaluation of the final model and for model training should be chosen carefully corresponding to the ecological objectives of the optimised model. The practical implications of these findings are clear. They provide further insight in the evaluation of ecological presence/absence models and attempt to assist modellers in their choice of suitable performance criteria.

The results of the theoretical analysis were confirmed in a first experimental analysis in this dissertation. A heuristic nearest ascent hill-climbing algorithm for rule base optimisation was applied to construct a fuzzy rule-based habitat suitability model for spawning European grayling (*Thymallus thymallus*) in the Aare river (Bern, Switzerland). Optimisation of the fuzzy rule-based model was based on two different training criteria, the weighted Correctly Classified Instances (CCI_w) and Cohen's *Kappa*. The ecological relevance of the results was assessed by comparing the optimised rule bases with a rule base derived from ecological expert knowledge. The results showed that model training based on CCI_w and based on *Kappa* lead to different final models and thus confirmed the theoretical analysis. Specifically, this dissertation showed that these different final models could be attributed to the inseparability of the data. If the data would have been separable, all optimisation criteria would have led to the same results. Optimisation based on *Kappa* appeared to generate acceptable results and was easier to use than optimisation based on CCI_w because the latter required fine tuning of a weight parameter which accounted for the species prevalence. However, this parameter provided more transparency in the model training process by quantifying the degree to which overprediction had to be stimulated to result in accurate models. Specifically, this approach allowed modellers to adjust the model training process to the model objectives and to gain insight into this process due to the quantification of the adjustment.

Although the CCI_w provided a first step towards adaptive model performance quantification, this performance criterion cannot deal with the fuzzy output of a fuzzy classifier because it is based on the confusion matrix. Such performance criteria, for instance, are not sensitive to the position of the fuzzy sets where the wrong classification occurs, and valuable information may thus be lost by transferring this fuzzy output to the crisp output which is needed to generate the confusion matrix. Therefore, the adjusted average deviation (*aAD*) was introduced. This performance measure includes a parameter α which ranges between 0 and 1 and allows stimulation of overprediction or underprediction, depending on the prevalence of the training data set.

To analyse the strengths and weaknesses of this approach, it was applied on model training and the results were compared to those obtained after model training based on some of the most frequently applied performance criteria: *Kappa*, *CCI*, the average deviation (*AD*) and the true skill statistic (*TSS*). The results suggest that *CCI*, *AD* and *TSS* criteria may be less appropriate for model training at extreme prevalence values. However, *TSS* and *Kappa* showed to be valuable criteria for the evaluation of the performance of the final model. Both accuracy-based measures, *CCI* and the *AD*, appeared to be less suitable for model evaluation because their values were correlated with the prevalence of the evaluation set. These findings confirm the results of the previous analysis on CCI_w and *Kappa* by showing that the performance criteria which are applied for model training may significantly affect the results of the training process.

Finally, this dissertation analysed the relation between the parameter α and the training set prevalence for three other species in different New Zealand river systems: the caddisflies *Aoteapsyche* spp., large brown trout *Salmo trutta* and rainbow trout *Oncorhynchus mykiss*. The *aAD* was implemented in a hill-climbing algorithm to optimise a fuzzy species distribution model for each species. For all studied species, the results proved that the optimal α value for model training depends on the prevalence of the training set. These results confirmed that *AD* and *CCI* may be less appropriate for model training due to their strong dependency on the training set prevalence. Moreover, these findings agreed that *Kappa* may be an appropriate performance criterion for robust model training. Model training based on *Kappa* led to results which were relatively accurate, whereas the results of model training based on *TSS* was more variable. Despite its user-friendliness, however, *Kappa* does not allow the model developers to focus on certain aspects of model performance and thus provides less control on the model training process than an adjustable performance criterion such as the *aAD*. Finally, this dissertation suggests that model training based on accuracy-based performance criteria may produce unrealistic models at extreme prevalences of the training dataset, whereas the *aAD* allows identification of more accurate and well-founded models. Moreover, comparison of the relation between the α parameter and the prevalence of the training data set for different species revealed this relation might be relatively general and may depend on data or model characteristics.

An optimal parameter value or training performance criterion could be found by applying sensitivity analysis, but a more important problem with these flexible performance criteria could be the difficulty to say which models are better. Although applying a training parameter which minimises the difference between over- and underprediction may be a suitable rule of thumb, there is no straightforward answer to this question, because the quality of a model depends on its purpose and the context in which the model will be applied. Conservationists should thus distinguish between models which reliably or less reliably predict species distribution. Given inseparable ecological data, each performance criterion focuses on different aspects of model performance, and these aspects are reflected in the final model which is obtained after training based on this performance criterion. Moreover, this dissertation showed theoretically that the aspects on which these criteria focus may vary with the prevalence of the training dataset. Consequently, conservationists and model developers should be aware of the different focus of the training performance criteria, and thus of the final model.

This dissertation may thus contribute to the transparency of the model training process and to an increasing awareness of the strengths and weaknesses of the final model. Specifically, the adjustable parameter in the *aAD* enables modellers to situate the optimised models in the heuristic search space and thus provides an indication of the ecological model relevance. Consequently, it may support modellers and river managers in the decision making process by enhancing model foundations. The habitat suitability models optimised in this dissertation are able to predict the effect of different impacts on the river system and to select the optimal restoration option. Hence, they could be a valuable decision support tool for river managers and ease the discussion between stakeholders. As such, they may be a small step on the road towards more reliable species distribution models. Due to the universality and the flexibility of the approach, it could be applied to any other ecosystem or species, and may therefore be valuable to ecological modelling and ecosystem management in general.

Samenvatting

Het rivierbeheer in Vlaanderen en de rest van de wereld was lange tijd gericht op de verbetering van de chemische waterkwaliteit, aan de hand van rioolwaterzuiveringsinstallaties of opgelegde effluentnormen. Deze ingrepen leidden wel tot een significante verbetering van de waterkwaliteit, maar desondanks is de fysische habitatkwaliteit vaak nog ontoereikend en stagneert de ecologische kwaliteit van vele rivieren. Kleinschalige ingrepen, zoals hermeandering, herstel van overstromingsgebieden en visdoorgangen zijn bijgevolg aangewezen om de streefdoelen van de Europese Kaderrichtlijn Water te halen en worden momenteel gepland of uitgevoerd in Europa. Om deze kleinschalige inspanningen efficiënt te verdelen is echter een grondige analyse vereist van de impact van deze ingrepen op de rivierecologie. Aangezien deze analyse en de identificatie van knelpunten in een rivierbekken kan gebeuren aan de hand van habitatgeschiktheids- of soortenverspreidingsmodellen, zijn deze modellen waardevolle instrumenten voor integraal rivierbeheer.

De structuur van dergelijke modellen moet ondubbelzinnig interpreteerbaar zijn zodat de gebruikers de onzekerheden van de modelresultaten eenvoudig kunnen nagaan. Vaaglogica is bijgevolg bijzonder geschikt om deze voorwaarden in te vullen. Deze theorie houdt enerzijds rekening met de inherente onzekerheid van de ecologische variabelen in het modelleringproces en biedt anderzijds een transparante weergave van niet lineaire verbanden tussen ecologische variabelen. Vage systemen kwantificeren variabelen aan de hand van linguïstische omschrijvingen zoals ‘laag’, ‘hoog’ of ‘matig’ en gebruiken ecologische expertkennis om deze omschrijvingen om te zetten in een mathematisch kader waarbinnen dataverwerking mogelijk is. Deze expertkennis is omvat in vage als-dan regels zoals ‘ALS de stroomsnelheid hoog is EN de diepte laag is DAN is de habitatgeschiktheid hoog voor een rheofiele vissoort’. Vage modellen gebruiken vaagverzamelingen om de onzekerheid in de omschrijving van de linguïstische waarden ‘hoog’, ‘laag’ en ‘hoog’ van de respectievelijke variabelen ‘stroomsnelheid’, ‘diepte’ en ‘habitatgeschiktheid’ te incorporeren in het model. Aangezien de grenzen tussen twee vaagverzamelingen overlappen, kan een object (bijv. een reëel getal) gedeeltelijk tot twee opeenvolgende verzamelingen behoren en kan een vage linguïstische omschrijving zoals ‘de diepte is redelijk laag maar eerder matig’ in het model worden verwerkt. Deze aanpak verschilt grondig met de klassieke verzamelingenleer waarin een object ofwel volledig ofwel niet tot een verzameling behoort.

Vage soortenverspreidingsmodellen kunnen ecologische aspecten, zoals de ecologische gradiënttheorie, implementeren in het modelleringproces. Aangezien deze aanpak de specifieke kenmerken van een ecologisch probleem weerspiegelt, zijn vage modellen bijzonder geschikt voor het modelleren van soortenverspreiding. De nood aan expertkennis is echter het grootste knelpunt bij de toepassing van vage modellen. Enerzijds is het vastleggen van relevante menselijke expertkennis vaak moeilijk en tijdrovend, terwijl anderzijds verschillende auteurs de consistentie van deze expertkennis betwisten. Recent onderzoek heeft aangetoond dat de combinatie van vage systemen met datagebaseerde technieken een oplossing kan bieden voor dit probleem. In het domein van de vage regelontwikkeling bijvoorbeeld worden vage regelgebaseerde modellen vaak opgesteld met behulp van heuristische zoekalgoritmes. Dit proefschrift wil een datagebaseerde optimalisatiemethode voor vage habitatgeschiktheidsmodellen ontwikkelen en testen. Hiervoor werd een hill-climbing algoritme toegepast om de vage expertregels van het model te optimaliseren, terwijl de impact van verschillende trainingscriteria op het optimalisatieresultaat werd geanalyseerd. De vaagverzamelingen werden geoptimaliseerd op basis van hun entropie, die de uniformiteit van de dataverdeling over de invoerruimte kwantificeert. De waarde van de voorgestelde methode voor integraal waterbeheer werd nagegaan aan de hand van 5 datasets die ecologische gevallenstudies beschrijven in België, Zwitserland en Nieuw-Zeeland.

De sterke en zwakke punten van een expertgebaseerde aanpak werden eerst geanalyseerd door vergelijking van deze aanpak met een datagebaseerde methode in een gevallenstudie in de Zwalm (Vlaanderen, België). Op basis van de beschikbare ecologische expertkennis en van biologische stalen uit het Zwalmbekken werden habitatmodellen opgesteld voor de eendagsvlieg *Baetis rhodani*, die een indicatorsoort is voor een goede ecologische kwaliteit. Zowel vage kennisgebaseerde als datagebaseerde habitatgeschiktheidsmodellen werden ontwikkeld en de performantie van deze modellen werd vergeleken. De datagebaseerde modellen steunden op een hill-climbing algoritme dat ecologische expertkennis afleidde uit de beschikbare data. Datagebaseerde modellen bleken significant beter te presteren dan de expertgebaseerde modellen, terwijl een step forward modelselectieprocedure aantoonde dat de habitatgeschiktheid voor de eendagsvlieg in het studiegebied het best werd verklaard door fysische habitatvariabelen zoals stroomsnelheid, rivierbreedte en het dominante rivierbodemsubstraat. Dit onderzoek heeft belangrijke gevolgen voor de toepassing van expertkennis in ecologische studies, vooral als deze kennis geëxtrapoleerd wordt naar andere gebieden. De resultaten suggereren dat datagebaseerde modellen kennisgebaseerde modellen kunnen aanvullen en zo de modelbetrouwbaarheid kunnen verbeteren.

Om betrouwbare datagebaseerde modellen te ontwikkelen is echter een correcte modeltraining en -evaluatie vereist. Het bepalen van de modelperformantie is een cruciale stap in deze procedures. Het performantiecriterium dat de modelperformantie kwantificeert, is de kern van deze performantiebepaling. Daarom geeft dit proefschrift een overzicht van de meest toegepaste performantiecriteria in de aan- en afwezigheidsmodellering van een soort. Bovendien werd de rol van deze criteria in de training en de evaluatie van het model geanalyseerd. Schatters van de voorspellingsperformantie worden algemeen aangewend om de kwaliteit van het finale model te bepalen, maar een uitgebreid literatuuroverzicht in dit proefschrift toonde aan dat de modeltraining meestal gebeurt met het oog op optimalisatie van de voorspellingsnauwkeurigheid van het model. Bovendien ontbreekt een systematisch overzicht omdat de meeste analyses van performantiecriteria empirisch werden uitgevoerd en zich enkel richten op specifieke kenmerken van de performantiecriteria. In dergelijk overzicht toonde dit proefschrift aan dat verschillende performantiecriteria een model (of zijn resulterende verwarringsmatrix) verschillend evalueren en dat dit verschil kan verklaard worden door de relatie tussen deze criteria en de prevalentie van de validatieset. Meer bepaald kan, bij de vergelijking van modellen afgeleid van verschillende datasets, de prevalentie van deze datasets de waarde van de performantiecriteria beïnvloeden en bijgevolg de resultaten van de vergelijking manipuleren. Gelijkaardige problemen kunnen optreden wanneer de performanties van een model worden vergeleken op basis van een trainingsset en een validatieset met verschillende prevalenties. Dit proefschrift toont theoretisch aan dat deze effecten zich enkel voordoen wanneer de data onscheidbaar zijn door een n-dimensioneel hypervlak, waarbij n het aantal invoervariabelen is. Door deze onscheidbaarheid richten verschillende performantiecriteria zich tijdens de modeltraining op verschillende aspecten van de modelperformantie, zoals sensitiviteit, specificiteit of voorspellingsaccuratesse. Deze bevindingen hebben belangrijke gevolgen voor het ecologisch modelleren aangezien ecologische data meestal onscheidbaar zijn door dataruis en de complexiteit van het bestudeerde systeem. Bijgevolg zouden gebruikers duidelijk moeten weten welk aspect van de modelperformantie gekwantificeerd wordt, en zouden modellen consistent moeten geëvalueerd worden, meer bepaald onafhankelijk van, of in functie van de prevalentie van de bestudeerde soort.

Dit proefschrift gaf ook aan dat de verschillende evaluatiescores van performantiecriteria voor hetzelfde model mogelijks veroorzaakt worden door het relatieve belang dat performantiecriteria hechten aan omissie- en commissiefouten. Een theoretische analyse leerde dat verschillende criteria een perfect model gelijk kunnen evalueren, maar tegelijkertijd toch kunnen focussen op sterk verschillende aspecten van de modelperformantie. Om een betrouwbare kwantificatie van de modelperformantie te garanderen, wordt het gebruik voorgesteld van criteria die onafhankelijk zijn van of rekening houden met de prevalentie. Bovendien zouden sommige criteria de modelontwikkelaars moeten toelaten om een onderscheid te maken tussen omissie- en commissiefouten, zowel tijdens training als tijdens modelevaluatie. De geschikte performantiecriteria voor de evaluatie van het finale model en voor de modeltraining zouden nauwgezet moeten gekozen worden in overeenstemming met de ecologische doelstellingen van het geoptimaliseerde model. De praktische implicaties van deze bevindingen zijn duidelijk. Ze verlenen dieper inzicht in de evaluatie van ecologische aan-/afwezigheidsmodellen en kunnen modelleerders helpen bij de keuze voor een geschikt performantiecriterium.

De resultaten van de theoretische analyse werden bevestigd in een eerste experimentele analyse in dit proefschrift. Een vaag regelgebaseerd habitatgeschiktheidsmodel voor kuitschietende Europese vlagzalm (*Thymallus thymallus*) in de Aare (Bern, Zwitserland) werd opgesteld met behulp van een heuristisch hill-climbing algoritme voor regelbankoptimalisatie. De optimalisatie van het vage regelgebaseerde model was gebaseerd op twee verschillende trainingscriteria: de gewogen Correctly Classified Instances (CCI_w) en Cohens *Kappa*. De ecologische relevantie van de resultaten werd bepaald door vergelijking van de geoptimaliseerde regelbank met een regelbank die werd afgeleid van ecologische expertkennis. De resultaten gaven aan dat modeltraining op basis van CCI_w en op basis van *Kappa* resulteert in verschillende finale modellen, wat de theoretische analyse bevestigde. Meer bepaald toonde dit proefschrift aan dat de onscheidbaarheid van de trainingsdata de geobserveerde verschillen veroorzaakt. Als de data scheidbaar waren, zouden alle optimalisatiecriteria tot hetzelfde resultaat geleid hebben. De optimalisatie op basis van *Kappa* bleek aanvaardbare resultaten op te leveren en was eenvoudiger toe te passen dan de optimalisatie gebaseerd op CCI_w omdat bij dit laatste criterium de gewichtsparameter, die de prevalentie in rekening bracht, moest worden afgestemd. Toch zorgde deze parameter voor meer duidelijkheid in de modeltraining aangezien deze de mate kwantificeerde waarin overpredictie moest gestimuleerd worden om tot accurate modellen te leiden. Meer bepaald liet deze aanpak modelleerders toe om het modeltrainingsproces aan te passen aan de doelstellingen van het model en om inzicht te krijgen in dit proces aan de hand van de kwantificatie van deze aanpassing.

De CCI_w vormt een eerste stap naar adaptieve kwantificatie van de modelperformantie, maar dit criterium kan de vage uitvoer van een vaag classificatiemodel niet verwerken omdat de CCI_w gebaseerd is op de verwarringsmatrix. Dergelijke performantiecriteria houden geen rekening met de positie van de vaagverzameling waar de foutieve classificatie plaatsvindt, en zo kan waardevolle informatie verloren gaan bij de omzetting van de vage uitvoer naar de scherpe uitvoer, die nodig is om de verwarringsmatrix op te stellen. Bijgevolg werd de aangepaste average deviation (*aAD*) ontwikkeld in dit proefschrift. Deze performantiemaat bevat een parameter α die varieert tussen 0 en 1 en die het stimuleren van overpredictie of onderpredictie toelaat, afhankelijk van de prevalentie van de trainingsdataset.

De sterke en zwakke punten van deze aanpak werden geanalyseerd door toepassing op modeltraining en vergelijking van de resultaten met deze die werden bekomen na modeltraining gebaseerd op een aantal algemeen toegepaste performantiecriteria: *Kappa*, *CCI*, de average deviation (*AD*) en de true skill statistic (*TSS*). De resultaten gaven aan dat *CCI*, *AD* en *TSS* minder geschikt zijn voor modeltraining bij uiterste prevalentiewaarden. *TSS* en *Kappa* bleken echter waardevolle criteria voor de evaluatie van de performantie van het finale model. De twee criteria die gericht zijn op nauwkeurigheid, *CCI* en *AD*, bleken minder geschikt te zijn voor modelevaluatie omdat hun waarden gecorreleerd waren met de prevalentie van de evaluatieset. De resultaten bevestigen dat de performantiecriteria voor modeltraining de resultaten van de modeltraining significant kunnen beïnvloeden en komen dus overeen met deze van de eerste analyse op basis van CCI_w en *Kappa*.

Ten slotte analyseerde dit proefschrift het verband tussen de parameter α en de prevalentie van de trainingsset voor drie andere soorten in verschillende riviersystemen in Nieuw-Zeeland: de kokerjuffers *Aoteapsyche* spp., grote beekforel *Salmo trutta* en regenboogforel *Oncorhynchus mykiss*. De *aAD* werd geïmplementeerd in een hill-climbing algoritme dat een vaag soortenverspreidingsmodel optimaliseerde voor elke soort. De resultaten toonden voor alle bestudeerde soorten aan dat de optimale α waarde voor modeltraining afhankelijk is van de prevalentie van de trainingsset. Deze resultaten bevestigden dat *AD* en *CCI* minder geschikt zijn voor modeltraining door hun afhankelijkheid van de prevalentie van de trainingsset. Bovendien bekrachtigen deze bevindingen dat *Kappa* mogelijks een geschikt performantiecriterium is voor een verkennende modeltraining. Modeltraining gebaseerd op *Kappa* resulteerde in relatief accurate resultaten, terwijl de resultaten van model training gebaseerd op *TSS* meer varieerden. Ondanks zijn gebruiksvriendelijkheid laat *Kappa* echter niet toe om te focussen op bepaalde aspecten van de modelperformantie en biedt bijgevolg minder controle over de modeltraining dan een aanpasbaar performantiecriterium zoals *aAD*. Dit proefschrift suggereert tenslotte dat modeltraining gericht op modelnauwkeurigheid kan leiden tot onrealistische modellen bij extreme prevalenties van de trainingsset, terwijl training op basis van de *aAD* kan resulteren in meer accurate en beter onderbouwde modellen. Een vergelijking van het verband tussen de parameter α en de prevalentie van de trainingsset voor de verschillende soorten in dit proefschrift wees bovendien uit dat dit verband mogelijks relatief algemeen is en afhankelijk van data- en modelkenmerken.

Een optimale parameterwaarde of een geschikt performantiecriterium kan gevonden worden aan de hand van sensitiviteitsanalyse. Hierbij kan het echter moeilijk zijn om aan te geven welke modellen beter zijn. Kiezen voor een parameterwaarde die het verschil tussen de over- en onderpredicties minimaliseert zou een gepaste vuistregel kunnen zijn, maar eigenlijk bestaat er geen eenduidig antwoord op deze vraag aangezien de kwaliteit van een model afhankelijk is van de doelstellingen en de context waarin het model zal worden toegepast. Beheerders zouden dus onderscheid moeten maken tussen soortenverspreidingsmodellen op basis van de betrouwbaarheid van hun voorspellingen. Gezien de onscheidbaarheid van de ecologische data, richt elk performantiecriterium zich op andere aspecten van de modelperformantie, en zijn deze aspecten weerspiegeld in het finale model dat bekomen wordt na training op basis van dit performantiecriterium. Bovendien toonde dit proefschrift theoretisch aan dat de aspecten waarop deze criteria zich richten kunnen variëren afhankelijk van de prevalentie van de trainingsdataset. Beheerders en modelontwikkelaars zouden bijgevolg rekening moeten houden met de verschillende focussen van de performantiecriteria voor modeltraining, en dus van het finale model.

Samengevat kan dit proefschrift bijdragen tot de transparantie van het modeltrainingsproces en tot een toenemende aandacht voor de sterke en zwakke punten van het finale model. De aanpasbare parameter in de *aAD* laat modelleerders toe om het geoptimaliseerde model te situeren in de heuristische zoekruimte en geeft een indicatie van de ecologische relevantie van het model. Door het versterken van de modelonderbouwing kan dit criterium bijgevolg modelleerders en rivierbeheerders helpen bij het beslissingsproces. De geoptimaliseerde habitatgeschiktheidsmodellen in dit proefschrift kunnen het effect van verschillende ingrepen op het riviersysteem voorspellen en zo de optimale hersteloptie selecteren. Op die manier kunnen ze bijgevolg een waardevol beslissingsondersteunend instrument vormen voor rivierbeheerders en kunnen ze de dialoog tussen verschillende belangengroepen vereenvoudigen. Dankzij het universele karakter en de flexibiliteit van deze modellen kunnen ze eenvoudig worden toegepast op andere soorten of ecosystemen, en zijn ze bijgevolg waardevol voor het ecologische modelleren en het ecosysteembeheer in de brede zin.

Curriculum vitae

Personalia

Naam	Ans Mouton
Geslacht	Mannelijk
Nationaliteit	Belg
Geboortedatum	29/03/1981, 26 jaar
Geboorteplaats	Assebroek, België
Burgerlijke stand	Ongehuwd
e-mail	Ans.Mouton@Ugent.be

Opleiding

1993 – 1999	Secundair onderwijs, Wiskunde-Wetenschappen (8u), Sint-Vincentiuscollege Eeklo
1999 – 2004	Bio-Ingenieur, optie Milieutechnologie, Universiteit Gent, Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen. *Scriptie*: Ontwikkeling van predictieve modellen voor het selecteren van herstelplannen in het Zwalmbekken. Promotoren: Prof. dr. N. De Pauw & Prof. dr. ir. P. Goethals
2006 – 2007	Academische Lerarenopleiding, optie Toegepaste Biologische Wetenschappen, grote onderscheiding. Universiteit Gent, Faculteit Psychologie en Pedagogische Wetenschappen
2004 – 2008	Doctoraatsopleiding in de Toegepaste Biologische Wetenschappen, Universiteit Gent, Faculteit Bio-ingenieurswetenschappen

Loopbaanoverzicht - werkervaring

2004 – 2008	Doctoraatsbursaal Bijzonder Onderzoeksfonds Universiteit Gent

Wetenschappelijke activiteiten

- ***Publicaties***

  - Internationale journal publicaties met peer review (ISI)

Mouton, A.M., Schneider, M., Depestele, J., Goethals, P.L.M., De Pauw, N. (2007). Fish habitat modelling as a tool for river management. Ecological Engineering 29, 305-315.

Mouton, A., Meixner, H., Goethals, P.L.M., De Pauw, N., Mader, H. (2007). Concept and application of the Usable Volume for modelling the physical habitat of riverine organisms. River Research and Applications 23, 1-14.

Mouton, A.M., Schneider, M., Armin, P., Holzer, G., Müller, R., Goethals, P.L.M., De Pauw, N. (in press). Optimisation of a fuzzy physical habitat model for spawning European grayling (Thymallus thymallus L.) in the Aare river (Thun, Switzerland). Ecological Modelling.

Mouton, A.M., Maes, A., Van Der Most, H., Jeuken, A., Goethals, P.L.M., De Pauw, N. (in press). Application of the Water Framework Directive Explorer in the Zwalm River Basin (Flanders, Belgium). River Research and Applications.

Mouton, A.M., Schneider, M., Kopecki, I., Wouters, H., Goethals, P.L.M., De Pauw, N. (submitted). Optimisation of physical habitat assessment for benthic fish and macroinvertebrates in lowland rivers. Journal of Environmental Monitoring.

Mouton, A.M., Goethals, P.L.M. (submitted). Towards biological valuation of freshwater ecosystems. Conservation Biology.

Dedecker, A.P., Mouton, A.M., Goethals, P.L.M.(submitted). Selecting variables for habitat suitability of *Asellus* (Crustacea, Isopoda) by applying input variable contribution methods to Artificial Neural Network models. Environmental Modelling and Assessment.

Mouton, A.M., De Baets, B. and Goethals, P.L.M. (submitted). Knowledge-based versus data-driven fuzzy habitat suitability models for river management. Environmental Modelling and Software.

Mouton, A.M., De Baets, B. and Goethals, P.L.M. (submitted). A theoretical analysis of performance criteria for the evaluation and optimisation of presence/absence species distribution models. Diversity and Distributions.

Mouton, A.M., De Baets, B., Van Broekhoven, E. and Goethals, P.L.M. (in preparation). Prevalence-adjusted optimisation of species distribution models. Ecological Modelling.

Mouton, A.M., Jowett, I.G., Goethals, P.L.M. and De Baets, B. (in preparation). Prevalence-adjusted optimisation of a fuzzy habitat suitability model for aquatic invertebrate and fish species in New Zealand. Journal of Fish Biology.

- Publicaties in internationale proceedings met peer review

Mouton, A.M., Schneider, M., Kopecki, I., Goethals, P.L.M., De Pauw, N. (2005). Application of MesoCASiMiR: Assessment of *Baetis rhodani* Habitat Suitability. In: Harby, A., Baptist, M., Duel, H., Dunbar, M., Goethals, P., Huusko, A., Ibbotson, A., Mader, H., Pedersen, M.L., Schmutz, S., Schneider, M. (Eds.), Proceedings of the COST 626 European Aquatic Modelling Network final meeting, 19-20 May 2005, Silkeborg, Denmark, p. 249-258.

Mouton, A.M., Schneider, M., Kopecki, I., Goethals, P.L.M., De Pauw, N. (2006). Application of MesoCASiMiR: Assessment of *Baetis rhodani* Habitat Suitability. In: Voinov, A., Jakeman, A.J., Rizzoli, A.E. (Eds.), Proceedings of the iEMSs Third Biennial Meeting: "Summit on Environmental Modelling and Software". International Environmental Modelling and Software Society, July 2006, Burlington, USA.

Mouton A.M., Schneider, M., Goethals P.L.M., De Pauw, N. (2006). Nature Knows – Heuristic Search Algorithms Optimizing a Fuzzy Rule-based Fish Habitat Suitability Model for Rivers. In: Proceedings of the ICEM 2006 International Conference on Ecological Modelling, 28 August-1 September 2006, Yamaguchi, Japan.

Mouton A.M., Maes A., Jeuken A., Van Der Most H., Goethals P.L.M., De Pauw, N. (2007). Application of the Water Framework Directive Explorer for river assessment and restoration in the Zwalm river basin (Flanders, Belgium). In: Biggs, B., Jowett, I., Hicks, B., Hayes, J., Nikora, V., Hicks, M., Larned, S., Coleman, S. (Eds.), Proceedings of the '6th International Symposium on Ecohydraulics: Bridging the gap between hydraulics and biology', 18-23 February 2007, Christchurch, New Zealand.

D'heygere, T., Adriaenssens, V., De Groote, D., Leyssen, A., Mouton, A., Carchon, P., Martens, K., Goethals, P.L.M., De Pauw, N. (2004). A priori and a posterio evaluation of river restoration projects by means of ecological models and biological monitoring: case studies in the Zwalm river basin (Belgium). In: García de Jalón, D. (Ed.), International Association of Hydraulic Research (IAHR), Proceedings of the '5th International Symposium on Ecohydraulics: Aquatic Habitats: Analysis & Restoration', 12-17 September 2004, Madrid, Spain.

Maddock, I., Dunbar, M., Eastman, K., Eisner, A., Hill, G., Mouton, A., Rivas-Casado, M., Smolar-Zvanut, N., Acreman, M. (2006). Assessing rapid physical habitat mapping methods for observer variability. In: Proceedings of the International Conference on "Riverine Hydroecology: advances in research and applications", 14-18 August 2006, Stirling, UK.

- Nationale journal publicaties met peer review

Maes, A., Mouton, A., Goethals, P.L.M., De Pauw, N. (submitted). KRW (Kaderrichtlijn Water)-Verkenner als instrument voor het ecologisch evalueren van rivierherstelopties. Water.

Mouton, A.M., Schneider, M., Peter, A., Holzer, G., Müller, R., Goethals, P.L.M., De Pauw, N. (submitted). Optimalisatie van een vaaglogisch fysisch habitatmodel voor kuitschietende Europese vlagzalm (Thymallus thymallus L.) in de Aare (Thun, Zwitserland)

- Andere journal publicaties

D'heygere, T., Mouton, A, Dedecker, A., Adriaenssens, V., Goethals, P.L.M., De Pauw, N. (2004). Development of predictive ecological models for river restoration. Communications in Agricultural and Applied Biological Sciences 69(2) 15-18.

Mouton, A., Dedecker, A., Stuer, V., De Ridder, K., Verbist, B., Adriaenssens, V., Goethals, P.L.M., De Pauw, N. (2005). Development of a biological water quality assessment system for the Sumberjaya watershed in Sumatra (Indonesia). Communications in Agricultural and Applied Biological Sciences 70(2) 199-202.

Mouton, A., Schneider, M., Peter, A., Holzer, G., Müller, R., Goethals, P., De Pauw, N. (2006). Optimization of a fuzzy physical habitat model for spawning European grayling (*Thymallus thymallus* L.). Communications in Agricultural and Applied Biological Sciences 71(1) 135-138.

Maes A., Goethals, P., Mouton A., De Pauw, N. (2006). The Water Framework Directive Explorer for decision support and integrated water management. Communications in Agricultural and Applied Biological Sciences 71(1) 179-183.

Mouton, A.M., Schneider, M., Depestele, J., Goethals, P.L.M., De Pauw, N. (2007). Simulation of the physical habitat suitability of bullhead in the Zwalm river. Communications in Agricultural and Applied Biological Sciences 72(1) 75-78.

- Hoofdstukken in boeken

Pahl-Wostl, C., Lindenschmidt, K.E., Janssen, J.A.E.B., Filatova, T., Mouton, A., Holtz, G., Van der Keur, P., Gaber, N. (in press). Complexity and uncertainty : rethinking the modelling activity. In : Jakeman, T., Voinov, A., Rizzoli, A., Chen, S. Integrated and participatory water resources management.

- ***Presentaties***

- Presentaties als spreker

D'heygere, T., Adriaenssens, V., De Groote, D., Leyssen, A., Mouton, A., Carchon, P., Martens, K., Goethals, P.L.M., De Pauw, N. A priori and a posterio evaluation of river restoration projects by means of ecological models and biological monitoring: case studies in the Zwalm river basin (Belgium). 5th International Symposium on Ecohydraulics: Aquatic Habitats: Analysis & Restoration, 12-17 September 2004, Madrid, Spanje.

Mouton, A.M., Dedecker, A. Water quality assessment in the Sumberjaya watershed (Sumatra, Indonesia). Workshop 'Water management in the Sumberjaya watershed (Sumatra, Indonesia)', 27 April 2004, ICRAF, Bogor, Indonesië.

Mouton, A.M., Schneider, M., Kopecki, I., Goethals, P.L.M., De Pauw, N. Application of MesoCASiMiR: Assessment of *Baetis rhodani* Habitat Suitability. COST 626 European Aquatic Modelling Network final meeting, 19-20 Mei 2005, Silkeborg, Denemarken.

Mouton, A.M., Schneider, M., Kopecki, I., Goethals, P.L.M., De Pauw, N. Application of MesoCASiMiR: Assessment of *Baetis rhodani* Habitat Suitability. iEMSs Third Biennial

Meeting: “Summit on Environmental Modelling and Software”. International Environmental Modelling and Software Society, Juli 2006, Burlington, VSA.

Mouton A.M., Schneider, M., Goethals P.L.M., De Pauw, N. Nature Knows – Heuristic Search Algorithms Optimizing a Fuzzy Rule-based Fish Habitat Suitability Model for Rivers. ICEM 2006 International Conference on Ecological Modelling, 28 Augustus-1 September 2006, Yamaguchi, Japan.

Mouton, A., Schneider, M., Depestele, J., Goethals, P.L.M., De Pauw, N. Kennisgebaseerde vaaglogische modellen voor vissen in stromende wateren. NecoV Wintersymposium, 17-19 Januari 2007, Nijmegen, Nederland.

Mouton A.M., Maes A., Jeuken A., Van Der Most H., Goethals P.L.M., De Pauw, N. Application of the Water Framework Directive Explorer for river assessment and restoration in the Zwalm river basin (Flanders, Belgium). 6th International Symposium on Ecohydraulics: Bridging the gap between hydraulics and biology, 18-23 Februari 2007, Christchurch, Nieuw Zeeland.

Mouton A.M., Maes A., Jeuken A., Van Der Most H., Goethals P.L.M., De Pauw, N. Toepassing van de KRW (Kaderrichtlijn Water) Verkenner voor de beoordeling en restoratie van rivieren in het Zwalmbekken. Workshop ‘Implementatie van de KRW-Verkenner in Vlaanderen en Nederland: resultaten en verder onderzoek. 20 Maart 2007, Gent, België.

Mouton A.M., Schneider, M., Depestele, J., Goethals, P.L.M. Simulation of the physical habitat suitability of bullhead in the Zwalm River. 13de PhD Symposium Faculteit Bio-ingenieurswetenschappen, 17 Oktober 2007, Leuven, België.

- <u>Presentaties als co-auteur</u>

Goethals, P., Mouton, A., Van Rolleghem, P., De Pauw, N. Data-driven and knowledge-based methods to develop predictive models for river communities. Symposium on Computational intelligence in water and environment. 15 December 2005, Delft, Nederland

Maddock, I., Dunbar, M., Eastman, K., Eisner, A., Hill, G., Mouton, A., Rivas-Casado, M., Smolar-Zvanut, N., Acreman, M. Assessing rapid physical habitat mapping methods for observer variability. International Conference on “Riverine Hydroecology: advances in research and applications”. 14-18 August 2006, Stirling, VK.

Maes, A., Goethals, P., Mouton, A., De Pauw, N. Implementatie KRW-Verkenner in Vlaanderen: Case study Burggravenstroom. Workshop ‘Implementatie van de KRW-Verkenner in Vlaanderen: Case-study Burggravenstroom. 10 November 2006, Gent, België.

Maes, A., Goethals, P., Mouton, A., De Pauw, N. Implementatie van de KRW-Verkenner in Vlaanderen: Case study Burggravenstroom. Workshop ‘Implementatie van de KRW-Verkenner in Vlaanderen en Nederland: resultaten en verder onderzoek. 20 Maart 2007, Gent, België.

- ***Poster presentaties***

Mouton, A., Dedecker, A., Stuer, V., De Ridder, K., Verbist, B., Adriaenssens, V., Goethals, P.L.M., De Pauw, N. Development of a biological water quality assessment system for the Sumberjaya watershed in Sumatra (Indonesia). Voorgesteld op:
-11th PhD Symposium on Applied Biological Sciences, 6 Oktober 2005, Leuven, België.
-B-IWA Happy Hour, 22 Maart 2005, Brussel, België.
-NecoV Wintersymposium, 17-18 Februari 2006, Brussel, België.
-Studiedag Starters Natuuronderzoek, 22 Maart 2007, Brussel, België.

Mouton, A., Depestele, J., D'heygere, T., Goethals, P., De Pauw, N. Development of ecosystem models to predict the effect of river restoration actions at different spatial scales. Voorgesteld op:
-NecoV Wintersymposium, 14-15 Januari 2004, Gent, België.

Depestele, J., Mouton, A., D'heygere, T., Goethals, P., De Pauw, N. Development of predictive models for the management of fish communities in the Zwalm River basin, Belgium. Voorgesteld op:
-B-IWA Happy Hour, 22 Maart 2004, Brussel, België.

Mouton, A., Schneider, M., Peter, A., Holzer, G., Müller, R., Goethals, P., De Pauw, N. Optimisation of a fuzzy physical habitat model for spawning European grayling (*Thymallus thymallus* L.). Voorgesteld op:
-Twelfth PhD Symposium on Applied Biological Sciences, 21 September 2006, Gent, België.
-NecoV Wintersymposium, 17-19 Januari 2007, Nijmegen, Nederland.

Maes, A., Goethals, P.L.M., Mouton, A., De Pauw, N. The Water Framework Directive Explorer for decision support in integrated water management. Voorgesteld op:
-Twelfth PhD Symposium on Applied Biological Sciences, 21 September 2006, Gent, België.
-B-IWA Happy Hour, 9 Oktober 2006, Brussel, België.

- ***Organisatie van wetenschappelijke meetings en conferenties***

Workshop 'Water management in the Sumberjaya watershed (Sumatra, Indonesia)', 27 April 2004, ICRAF, Bogor, Indonesië. Organisatoren: Verbist B., Dedecker A., Mouton A.

WFD-Explorer workshop: Toepassing van de KRW-Verkenner – pilotstudy Burggravenstroom. 19 Mei 2006, Universiteit Gent, Gent. Organisatoren: Maes A., Mouton A., Goethals P.

WFD-Explorer workshop: Presentatie van de KRW-Verkenner aan Vlaamse waterbeheerders. 10 November 2006, Universiteit Gent, Gent. Organisatoren: Maes A., Mouton A., Goethals P.

WFD-Explorer workshop: Implementatie van de KRW-Verkenner in Vlaanderen en Nederland: resultaten en verder onderzoek. 20 Maart 2007, Gent. Organisatoren: Maes A., Mouton A., Goethals P.

Onderwijsactiviteiten

- ***Practica aan de Universiteit Gent***

2004-heden: Biological Monitoring and Water Quality Assessment (Prof. dr. N. De Pauw): Praktische oefeningen voor MSc. Environmental Sanitation and Technology, Engelstalig programma (15h).

2004-heden: Biologische Waterkwaliteitsbeoordeling (Prof. dr. N. De Pauw): Praktische oefeningen voor Bio-Ingenieurs (optie Milieutechnologie) (15h).

2004-heden: Biologische Waterkwaliteitsbeoordeling (Prof. dr. N. De Pauw): Praktische oefeningen voor MSc. Milieusanering en -technologie, Nederlandstalig programma (15h).

- ***Begeleiding van scripties en stages***

BSc. Meirlaen Tim (2004-2005). Stage & Scriptie. Microhabitatstudie van benthische macro-invertebraten in het Zwalmbekken. Katholieke Hogeschool Sint-Lieven, Departement Gent, Graduaat Chemie, optie Milieuzorg. Promotoren: N. De Pauw, K. Vossen. Begeleider: A. Mouton.

ir. Taelman Elisa (2004-2005). Scriptie. Ontwikkeling en toepassing van ecologische modellen voor macro-invertebraten aan de hand van expertkennis. Universiteit Gent, Bio-ingenieur in de Milieutechnologie. Promotoren: N. De Pauw, P. Goethals. Begeleider: A. Mouton.

ir. Anthierens Annelies (2004-2005). Scriptie. Uitwerken van herstelopties binnen het Zwalmbekken in het kader van geïntegreerd rivierbeheer. Universiteit Gent, Bio-ingenieur in het Land- en Bosbeheer. Promotoren: N. De Pauw, P. Goethals. Begeleider: A. Mouton.

ir. Smet Johan (2004-2005). Scriptie. Analyse en optimalisatie van habitatmonitoring voor macro-invertebraten in rivieren. Universiteit Gent, Bio-ingenieur in het Land- en Bosbeheer. Promotoren: N. De Pauw, P. Goethals. Begeleider: A. Mouton.

MSc. Lammertyn Matthew (2005-2006). Scriptie. Optimalisatie van nutriëntenbeheer in tropische aquaria via een combinatie van microbiële gemeenschappen, planten en macro-invertebraten. Universiteit Gent, Master na Master in de Milieusanering en het Milieubeheer. Promotoren: W. Verstraete, P. Goethals. Begeleider: P. Deschrijver, A. Mouton.

ir. Michiels Pol (2005-2006). Scriptie. Habitatanalyse van macro-invertebraten op basis van laboratoriumexperimenten. Universiteit Gent, Master na Master in de Milieusanering en het Milieubeheer. Promotoren: N. De Pauw, P. Goethals. Begeleider: A. Mouton.

MSc. Kwassenwi Bakari Collette (2005-2006). Scriptie. Ecological impact assessment of dams and weirs. Ghent University, Master of Science in Environmental Sanitation. Promotoren: N. De Pauw, P. Goethals. Begeleider: A. Mouton.

ir. Quintelier Els (2006-2007). Scriptie. Optimalisatie van gegevensgebaseerde kennisontwikkeling voor ondersteuning van rivierbeheer. Promotoren: P. Goethals, A. Mouton. Begeleider: A. Mouton.

ir. Monseré Tine (2006-2007). Scriptie. Koppeling van habitatgeschiktheid en biologische waardering van rivieren in Vlaanderen. Universiteit Gent, Bio-ingenieur in het Land- en Bosbeheer. Promotoren: P. Goethals, A. Mouton. Begeleider: A. Mouton.

Lic. Desomere Piet (2007-2008). Scriptie. Analyse van de impact van DDT en derivaten op benthische invertebraten in Vlaanderen. Universiteit Gent, Master na Master in de Milieusanering en het Milieubeheer. Promotoren: P. Goethals, A. Mouton. Begeleider: A. Mouton.

Lic. Esther Castermans (2007-2008). Scriptie. Modelgebaseerde habitatanalyse van steenvliegen (Plecoptera) in rivieren. Universiteit Gent, Master na Master in de Milieusanering en het Milieubeheer. Promotoren: P. Goethals, A. Mouton. Begeleider: K. Lock.

Wetenschappelijke onderscheidingen en beurzen

- ***Wetenschappelijke onderscheidingen***

Environmental Modelling and Software Society (EMSS), Best paper award for the best student paper and presentation at the iEMSs Third Biennial Meeting: “Summit on Environmental Modelling and Software”. International Environmental Modelling and Software Society, Juli 2006, Burlington, VSA, op basis van de paper en presentatie ‘Application of MesoCASiMiR: Assessment of *Baetis rhodani* Habitat Suitability’. Mouton, A.M., Schneider, M., Kopecki, I., Goethals, P.L.M., De Pauw, N.

Nederlands-Vlaamse Vereniging voor Ecologie (NecoV), Poster award, Wintermeeting 14-15 Januari 2004, Gent, Belgium, op basis van de poster ‘Development of ecosystem models to predict the effects of river restoration projects at different spatial scales’. Mouton, A., Depestele, J., D’heygere, T., Goethals, P.L.M., De Pauw, N.

- ***Onderzoeksbeurzen voor buitenlandse verblijven (* = langere duur)***

COST 626 workshop on the assessment of rapid physical habitat mapping methods for observer variability. 26 Juli – 6 Augustus 2004, CEH, Wallingford, UK. Contactpersoon: dr. Michael Dunbar.

COST 626 meeting on FST measurements and shear stress. 8-13 December 2004, NERI, Silkeborg, Denemarken. Contactpersoon: dr. Nicolas Friberg.

Vlaamse Interuniversitaire Raad (VLIR), Zuidinitiatief in het kader van het Eigen Initiatieven Programma 2005, Evaluation, optimisation and implementation of biological monitoring techiques to improve the Way Besai river water quality in Indonesia. Maart-April 2005*, International Centre for Research on Agroforestry (ICRAF) en University of Bogor, Bogor, Indonesië. Contactpersoon: ir. Bruno Verbist, KUL.

COST 626 Short Time Scientific Mission, 20 Februari- 5 Maart 2005, Stuttgart University en sje Schneider & Jorde Ecological Engineering, Stuttgart, Duitsland. Contactpersoon: dr. Matthias Schneider, sje, Universiteit Stuttgart.

Scientific Mission, November-December 2005*, Stuttgart University en sje Schneider & Jorde Ecological Engineering, Stuttgart, Duitsland. Contactpersoon: dr. Matthias Schneider, sje, Universiteit Stuttgart.

Scientific Mission, Februari-Maart 2007*, National Institute of Water and Atmospheric Research, Hamilton, Nieuw Zeeland. Contactpersoon: dr. Ian Jowett, NIWA.